AN EXAMINATION OF THE PRACTICE OF CULTURALLY RESPONSIVE EVALUATION: A CASE STUDY OF THE GIRLS ADVENTURES IN MATHEMATICS, ENGINEERING, AND SCIENCE (G.A.M.E.S.) SUMMER CAMP AT THE UNIVERSITY OF ILLINOIS

BY

MARIA BLANCA JIMENEZ

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Educational Policy Studies
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Professor James Anderson, Chair
Professor Lizanne DeStefano
Professor Stafford Hood
Professor William Trent

Abstract

In recent decades, there has been an expansion of evaluation approaches that seek to respond to the cultural context of the program that is being evaluated. Of these, culturally responsive evaluation is the most well-known evaluation approach. The purpose of culturally responsive evaluation is to take into account the cultural context of the program in the evaluation. To do so, it incorporates the culture of the program and participants throughout the evaluation process: preparation, design, and implementation. This instrumental case study examines how a culturally responsive evaluation approach was applied to evaluate a Science, Technology, Engineering, and Mathematics (STEM) program. More specifically, it highlights the specific culturally responsive evaluation strategies that were employed throughout the evaluation process. Further, it highlights the strengths and challenges associated with a culturally responsive evaluation approach.

The results of the present study support the future use of culturally responsive evaluation approach. While challenges common to culturally responsive evaluation were evident in the study, the strengths were certainly outweighed by the challenges. First, stakeholder engagement increased the overall credibility of the evaluation. For example, the input and advice of the program staff on the evaluation increased the quality of the design of the evaluation and generated evaluation results that were useful and meaningful to program staff and participants. Secondly, a culturally diverse evaluation team increased the validity of the evaluation through the use of shared lived experiences which allowed for the accurate collection, analysis, and interpretation of evaluation findings. Thirdly, a robust evaluation design (mixed methods) that was inclusive of culturally responsive evaluation instruments, analysis, and interpretation also increased the validity of the overall evaluation and results.

Further, analysis of data from a culturally responsive evaluation lens provided for the accurate depiction of the experiences of the participants, and in some cases revealed cultural differences that were addressed important to the evaluation's purpose. The present study is significant in that it adds to the current body of literature on culturally responsive evaluation. In addition, it calls for additional strategies and guidelines for conducting culturally responsive evaluation strategies in order to increase the practice of culturally responsive evaluation both domestically and internationally.

Dedicated to my late father and mother Daniel Jimenez and Blanca Jimenez, and to my fellow Latinos, Si Se Pudo! Thank you for your inspiration, love, encouragement, support, hard work, and sacrifice.

Acknowledgments

First, I would like to thank my committee members for your countless support, advice and mentorship. Dr. James Anderson: I admire you and because of you I came to the UIUC. To Dr. Lizanne DeStefano: You are a role model, thank you for all the opportunities that you have given me. Dr. William Trent: Thank you for feedback on drafts of my dissertation, I have learned a lot from our conservations. Dr. Stafford Hood: You are also a role model and you inspire me to continue to work in the field of evaluation.

My sincere gratitude to my second home and family: I-STEM. Linda, you are my mom away from home. Ayesha, without you the last couple of years of my graduate career would have not been as fruitful, Gabi, you are a great amiga, thank you, keep your head up, Holly, we shared great moments, and you make me stronger, Betsy, thank you for all of your editorial help and for bringing a smile to my face, Trinidad Molina, you are one of my dearest friends, Evelyn Perez, you make me laugh, New Orleans forever, Joey Perez, you are a trailblazer, Victor Perez, John Arrona, my fellow Slug, Casey George-Jackson, Erin Castro, and many many more, thank you for your countless years of support and encouragement. A special thanks to my partner-in-crime: Xochitl Casillas, we have been through a lot over the years but you have always been by my side. To my girlfriends in California, Lourdes Madrigal and Angie Gonzalez, thank you for your support, true friends always pick up where they left off. To my family, Mami, *tu eres mi inspiration y razon de vivir*, to my brother Victor Serrano, thank you for checking up on me from time to time, to my sister Sandra Pena, thank you for being the best sister in the world. To Mi Amor, Edgar Ramirez, thank you for sharing your life with me. And most of all thank you Dad, for teaching me the value of an education and for preparing me for the world, without you I would have never made it through this challenging journey. I dedicate this to you!

Table of Contents

List of Tables vii

List of Figures viii

Chapter 1 Introduction 1

Chapter 2 Literature Review 11

Chapter 3 Methodology 61

Chapter 4 Evaluation Context 79

Chapter 5 Findings: Preparing For, Designing, and Implementing Culturally Responsive Evaluation 103

Chapter 6 Findings: Strengths and Challenges 136

Chapter 7 Findings: Discussion and Recommendations 157

References 171

Appendix A The Educative, Values Engaged Framework 178

Appendix B Evaluation Presentation Slides 179

Appendix C Observation Guide 180

Appendix D Evaluator Interview Protocol 182

Appendix E Program Administrator Interview Protocol 184

Appendix F Rubric 186

Appendix G Sample Materials 188

Appendix H Sample Day Schedule for G.A.M.E.S. 191

Appendix I Participant Survey 192

Appendix J Focus Group Interview Protocol 201

Appendix K Parent Interview Protocol 203

Appendix L Camp Counselor Interview Protocol 205

List of Tables

Table Page

1 *Data Collection Process* 73

2 *Percent of Participants by Camp in 2007* 88

3 *Percent of Participants by Camp in 2008* 88

4 *Percent of Participants by Camp in 2009* 89

5 *Race/Ethnicity of Participants in 2007* 89

6 *Race/Ethnicity of Participants in 2008* 90

7 *Race/Ethnicity of Participants in 2009* 90

8 *Comparison of the G.A.M.E.S. Camps and NSES* 98

9 *Key Evaluation Questions and Methodology* 102

10 *Evaluation Question 1 and Methodology* 117

List of Figures

Figure Page

1 Culturally responsive evaluation framework 44

2 Sample questions from the participant survey 129

3 Pretest-posttest comparisons of survey results for all campers 130

4 Average response to survey question 4 disaggregated by race and camp 132

CHAPTER 1

Introduction

Culturally Responsive Evaluation

The field of program evaluation provides evaluators with a range of approaches from which to choose when designing and implementing program evaluations. In recent decades, a proliferation of evaluation approaches that account for culture and context have emerged. In addition to traditional evaluation approaches, researchers now choose from a wide range of evaluation approaches that attempt to be responsive to the culture and context of the program being evaluated. Examples of evaluation approaches which promote cultural sensitivity and responsiveness include: culturally competent (SenGupta, Hopson, & Thompson-Robinson, 2004), culturally responsive (Hood, Hopson, & Frierson, 2005), inclusive (Mertens, 1999; 2003), empowerment (Fetterman, 1994), and cross-cultural (Chouinard & Cousins, 2009). Similar in their aim, these evaluation approaches seek to value and respect the cultural context of the program in the evaluation.

Of the evaluation approaches which emphasize the role of culture in evaluation, perhaps the most well known is culturally responsive evaluation. Culturally responsive evaluation is similar to other evaluation approaches which are based on a "set of values or ideology" (Ryan, Chandler, & Samuels, 2007, p. 201). Moreover, culturally responsive evaluation extends the major tenets of other evaluation approaches including responsive evaluation (Stake, 1975; Abma & Stake, 2001; House, 2001; Stake, 1995; 2003) and participatory evaluation (Cousins & Whitmore, 1998; King, 1998; 2007). For example, culturally responsive evaluation responds to the evaluation's requirement for feedback, a major tenet of responsive evaluation. Further, culturally responsive evaluation incorporates the main tenet of participatory evaluation, which is stakeholder engagement. For example, culturally responsive evaluation outlines specific roles

that the stakeholders can play in the evaluation (i.e. team players, designers, and implementers of the evaluation) and evaluation stakeholders throughout the process (Ryan et al., 2007).

The primary goal of culturally responsive evaluation is to respond to cultural factors present within the evaluation context. Further, culturally responsive evaluation fully considers the culture and context of the program being evaluated. In addition, it responds to the culture of the participants and fully engages with the values of the program. To design a culturally responsive evaluation, various components of the program's culture and context are embedded within the evaluation. More specifically, the culture and context of the program is embedded at every phase of the evaluation, for example, in the design of the evaluation questions to the implementation of the evaluation results. Further, strategies that are culturally aligned with the context of the program are employed to design and implement the evaluation. Culturally responsive evaluation strategies respect, are sensitive to, and respond to the cultural context of the program and its participants (Frierson, Hood, & Hughes, 2002; Frierson, Hood, Hughes, & Thomas, 2010; Manswell-Butty, Reid, & LaPoint, 2004).

In addition to incorporating culture and context in the evaluation, culturally responsive evaluation also seeks to honor the cultural context of the program (Frierson, et al., 2002; Ryan et al., 2007; Frierson, et al., 2010). Culturally responsive evaluation advocates for the use of culturally diverse evaluation teams. From a culturally responsive evaluation approach, minority evaluators possess shared lived experiences which enable them to better assess, interpret, and understand the cultural context of the program. Minority evaluators can also better relate to the culture of the participants in the evaluation. Further, culturally responsive evaluators self-reflect on their cultural values in relation to the context of the evaluation (Ryan et al., 2007). Finally, culturally responsive evaluation evaluators seek to understand the program in the evaluation

within its own circumstances and counter any assumptions or judgments about the cultural context based on their own personal beliefs or values.

Main Tenets

The three main components of culturally responsive evaluation are: (a) context, (b) culture, and (c) responsive evaluation (Manswell-Butty, et al., 2004; Thomas, 2004; Ryan, et al., 2007). These three elements are incorporated at every phase of the evaluation, in the preparation, design, and implementation of the evaluation. Further, concepts of culture and context are embedded in the following stages of the evaluation process: 1) prepare for the evaluation, 2) engage stakeholders, 3) identify the purpose of the evaluation, 4) frame the right questions, 5) design the evaluation, 6) select and adapt instrumentation, 7) collect the data, 8) analyze the data, and; 9) disseminate and use the results (Hopson, 2010). Below is a description of each component of culturally responsive evaluation.

Context. The first component is context. Context includes factors present within the evaluation context which can have an impact on the results of the evaluation (Butty, Reid, & LaPoint, 2004). The context of an evaluation includes the "totality of environment in which the project takes place" (Manswell-Butty et al., 2004, p. 38). Factors which are related to the context of the program include the: (a) geographical location, (b) timing, (c) political and social environment, and (d) economic conditions (Manswell-Butty et al., 2004; Thomas, 2004). In addition, context also includes the: (a) material conditions of the setting; (b) institutional and organizational climate of the funding agency; (c) personal and interpersonal relationships between the participants; and (d) political dynamics of the program, such as power, privilege, and influence (Greene, 2005).

Culture. The second component is culture. There are many different ways to define culture (Manswell-Buddy et al., 2004; Ryan et al., 2007). In culturally responsive evaluation, the concept is defined broadly. First, culture includes the learned and shared behaviors, values, customs, and beliefs that influence and shape human behaviors, such as norms, customs, spiritual and religious practices, language, and educational systems (Frierson, et al., 2002, Butty, et al., 2004; SenGupta, et al., 2004; Frierson, et al., 2010; Kirkhart, 2010; Samuels & Ryan, 2011). Second, culture includes the social and economic status, nationality, religion, race, ethnicity, and primary language of a particular group in society (McDonald, Kutura, Richmond, & Betts, 2004). Lastly, culture includes the historical, environmental, economic, and political aspects which have an impact on how groups or individuals live and interact with other (Botcheva, Shih, & Huffman, 2009).

In culturally responsive evaluation, evaluations are designed based on the notion that culture is fluid, never static, and evolves over time from one group to the next (Kirkhart, 2010). That is, culture is transmitted from one generation to the next, and is the never the same from context to context. In addition to fluidity, culturally responsive evaluation acknowledges that culture is also never neutral and exists at multiple levels (Kirkhart, 2010). For example, culture is embedded within a specific set of historical, social, political, and economic contexts (Botcheva, et al., 2009). In addition, issues of power and privilege influences how programs, institutions, and organizations operate. Lastly, culture can exist at multiple levels: individual and institutional level. Individual levels of culture refer to the ways in which individual's identify with "categorical labels attached to personal characteristics such as race, ethnicity, religion, and gender" (Kirthart, 2010, p. 401). Culture at the institutional level refers to the ways in which

institutions impart "knowledge, values, beliefs, and skills that are also shared and communicated across cohorts or generations" (p. 401).

Responsive evaluation. The third dimension is responsive evaluation. Responsive evaluation is rooted in Robert E. Stake's work and is an evaluation approach which responds to the need for evaluative feedback (Stake, 1975; 2003; 2005). To design a responsive evaluation, evaluators spend a significant amount of time in the evaluation context observing program activities and identifying key issues of interest to the program staff. Responsive evaluation does not adopt a predetermined set of goals for the evaluation; rather, it addresses issues of importance to the stakeholders as they arise within the program context. From this perspective, the evaluation responds to issues which are not only important to the evaluation but are valued by the different stakeholders. Therefore, responsive evaluation provides useful information for program improvement, as well as providing space for the values/perspectives of the evaluation participants (i.e. program staff, administrators, participants, etc.) to surface within the evaluation (Stake, 1995; House, 2001; Butty et. al, 2004).

Research Problem

Recent studies have documented the benefits of culturally responsive evaluation (LaPoint & Jackson, 2004; Manswell-Butty, et al., 2004; Guzman, 2003; Ryan, et al., 2007), including better documentation of program outcomes, greater use of the evaluation results, and added credibility of the evaluation (Manswell-Butty, et al., 2004; Frazier-Anderson, Hood, & Hopson, 2011; Askew, Beverly, & Jay, 2010). Despite the known benefits of culturally responsive evaluation, there are some scholars in the field who do not support the use of culturally responsive evaluation (Frierson, et al. 2002; Frierson, et al., 2010). More specifically, there are two main arguments that have been put forth against the use culturally responsive evaluation.

The first is that evaluation should be culture-free, and the second is that evaluators should not be responsive to the culture and values of the program being evaluated (Frierson, et al., 2002; Frierson, et al., 2010). In many ways, these claims against culturally responsive evaluation stem from a lack of knowledge about the use and implications of culturally responsive evaluation.

In a recent article written by Chouinard & Cousins (2009), the authors noted several gaps in the field related to the practice of culturally responsive evaluation. They included: 1) a lack of knowledge on how to incorporate concepts of culture and context into evaluation practice, and 2) a lack of knowledge on how to conduct and implement culturally responsive evaluations in culturally diverse evaluation contexts. In addition to gaps in the literature, some scholars have questioned the validity and worth of culturally responsive evaluation. For example, some scholars have posed the question: 1) How useful and valid are identified program outcomes? (Manswell-Butty, et al., 2004). Further, a number of studies have argued for a clearer and more concise conceptualization of the practice of culturally responsive evaluation. For example, some scholars have posed questions which include: 1) How can staff design evaluations to ensure that all stakeholders feel empowered and part of the evaluation process?, or 2) How can evaluators include the context and culture of the program participants during the evaluation process (Manswell-Butty, et al., 2004; Hood, et al., 2005; Chouinard & Cousins, 2009).

The gaps in the literature regarding the practice of culturally responsive evaluation can be attributed to a number of factors. First, culturally responsive evaluation is a relatively new evaluation approach. Therefore, "little research has been conducted to examine how newer evaluation paradigms, such as culturally responsive evaluation, have defined, assessed, and interpreted program outcomes and impact" (Manswell-Butty, et al., 2004, p. 38). Two, there are only a limited number examples of culturally responsive evaluations in the field which

demonstrate the ways in which culture and context are fully incorporated throughout the evaluation process. Because culturally responsive evaluations require a lot of time, funding, and materials to conduct properly, they do not always fulfill their full potential. Thirdly, limited research has addressed *how* culturally responsive evaluation is prepared, designed, and implemented. For example, the evaluation designs, methods, and instruments that can be employed to include culture and context in a culturally responsive evaluation have not been adequately addressed in the literature. Therefore, this research study seeks to address this void in the literature by exploring how evaluators develop and implement culturally responsive evaluations.

This research project is in accordance with various efforts pursued by current scholars in the field to promote greater use of culturally responsive evaluation. For example, Rodney Hopson (2010) and others have developed a practical framework for conducting culturally responsive evaluations. In addition, two guidebooks on the topic evaluations have been published, one in 2002 and a revised version in 2010 by the National Science Foundation (NSF) (Frierson, et al., 2002; Frierson, et al., 2010). Both guidebooks were written by prominent scholars in the field and provide a multitude of suggestions and strategies for developing culturally responsive evaluations. Additionally, more and more evaluators have published articles, book, and excerpts on culturally responsive evaluation. Some of these texts offer examples and lessons learned regarding this particular approach to culturally responsive evaluation (Manswell-Butty, et al., 2004; Ryan, et al., 2007). Despite these efforts, there remain some limitations in our knowledge regarding the uses and implications of culturally responsive evaluation. Moreover, improving evaluation requires a clearer understanding of culturally responsive evaluation, more specifically the benefits and/or challenges to conducting this type of

evaluation, in order to further expand the use of this and similar evaluation approaches which necessitate the importance of culture and context in the evaluation.

Purpose of the Study

This research project is concerned with the practice of culturally responsive evaluation. The objective of this research project is to develop a more concise and comprehensive understanding of the uses and implications of culturally responsive evaluation. Toward this end, this study will present a case example of a culturally responsive evaluation in the form of a case study. This case study will investigate how culturally responsive evaluation was applied in an evaluation of a Science, Technology, Engineering, and Mathematics (STEM) program. The evaluation under consideration in this case study is the Girls' Adventures in Science, Mathematics, Engineering, and Technology (G.A.M.E.S.) summer camp located in the College of Engineering at the University of Illinois at Urbana-Champaign (UIUC). A more comprehensive description of this program will be provided in Chapter 3.

In an attempt to examine the practice of culturally responsive evaluation, this study took the following steps. First, this case study explored the rationale for the use of culturally responsive evaluation in the evaluation. Second, it examined the ways in which culture and context were incorporated at every aspect of the evaluation phases. Third, it identified strategies which were used to develop and implement a culturally responsive evaluation. Thirdly, it highlighted various strengths and challenges of conducting culturally responsive evaluation. In the end, this study aimed to promote a more comprehensive understanding of the practice culturally responsive evaluation.

In order to achieve the research objectives previously mentioned, the main research question in this case study is: *"How do evaluators plan, design, and implement culturally responsive evaluations?"*

This case study also addressed subsidiary questions whose answers helped to inform the main research question. Subsidiary questions include:

1. What was the rationale for using a culturally responsive evaluation approach?
2. What strategies were used for incorporating the cultural context of the program in the evaluation?
3. What strategies were used for incorporating the culture of the participants in the evaluation?
4. What strengths and challenges did the evaluators face in conducting a culturally responsive evaluation?

Significance of the Study

Presently, there is a need to further examine the practice of culturally responsive evaluation. To date, there is very little research which outlines what culturally responsive evaluation looks like in practice, and how evaluators can incorporate dimensions of the cultural context of the program in the evaluation. In addition, the field needs to expand beyond an advocacy of why culture is important to take into account in the evaluation, but address "the range of evidence available of throughout nontraditional evaluation models" (Masnwell-Butty, 2004, p. 38). Therefore, a closer examination of the uses and implications of this evaluation approach would help render important information needed for advancements in the field. To do so, the field calls for more examples of culturally responsive evaluation to help inform and

develop a better understanding of the practice of culturally responsive evaluation, which will ultimately increase our knowledge of its theoretical implications in the field.

The current state of the literature gives this research project significance in a number of ways. First, the study extends the current body of literature on culturally responsive evaluation which is limited in terms of the number of examples of culturally responsive evaluation. Second, it examines how culture and context were incorporated in an evaluation. Thirdly, it identifies key evaluation strategies which were employed to design and implement the evaluation. Finally, the study explores the strengths and limitations of conducting culturally responsive evaluation and offers a list of recommendations based on the lessons learned in the evaluation. Overall, the study informs key questions in the field which have yet been adequately addressed. For example, How do evaluators include the context and culture of the program participants during the evaluation? or How useful or valid are program outcomes in culturally responsive evaluation? (Manswell-Butty, et al, 2004, p. 38). Therefore, this study helps to inform the body of literature on the practice of culturally responsive evaluation in multiple ways.

The following chapter provides a review of literature on culturally responsive evaluation. The chapter will lay the theoretical framework for the study by outlining the major components of culturally responsive evaluation. Additionally, this chapter will discuss the origins of culturally responsive evaluation as well as contemporary issues of interest within the field.

CHAPTER 2

Literature Review

This chapter summarizes the literature on culturally responsive evaluation. First, it provides an overview of culturally responsive evaluation. Second, it provides an analysis of the roots of culturally responsive evaluation. Third, it examines responses to culturally responsive evaluation in the field of evaluation. Fourth, two practical frameworks for conducting culturally responsive evaluation are presented. Lastly, the benefits and challenges of conducting culturally responsive evaluation are explored. The chapter ends with future considerations for the field of culturally responsive evaluation.

Culture in Evaluation

In the past decade, the role of culture has gained attention in the field of evaluation (Hopson, 2003; Hood, et al., 2005). The increased role of culture in the evaluation field can be attributed to a number of factors. Firstly, the growing diversity of the nation's population has "created a heightened awareness of and need for evaluators to understand the complexity and multidimensionality of evaluations within multicultural, multiracial, and cross-cultural contexts" (Samuels & Ryan, 2011, p. 183). Secondly, paradigm shifts in the field, particularly from post-positivism to interpretivism and critical race theory, have created spaces for evaluation frameworks which incorporates concepts of culture and context have emerged (Manswell-Butty, et al., 2004; Chouinard & Cousins, 2009). Thirdly, an increase in evaluators of color with a commitment to social justice in evaluation has also lead to the increased attention of the evaluation field on the role of culture in evaluation. More specifically, a growing presence of evaluators of color (i.e. Rodney Hopson, Stafford Hood, Veronica Thomas, Ricardo Millett,

Melvin Hall, Saumitra SenGupta, and Henry F. Frierson)[1] over the past decade has provided the opportunity for conversations on culture in evaluation to take place.

Today, the importance of culture in the evaluation field can be seen in various places. For example, the presence of culture in evaluation is evident in the American Evaluation Association's (AEA) Guiding Principles. The 2004 AEA Guiding Principles for Evaluators on cultural competence purports that "evaluators should possess (or ensure that the evaluation team possesses) the education, abilities, skills and experience appropriate to undertake the tasks proposed in the evaluation" (American Evaluation Association, 2004). Furthermore, the principles require that evaluators are reflective about their own cultural assumptions and biases and develop evaluations that are respectful and appreciative of the program context (AEA, 2004; Askew, et al., 2012). Moreover, the presence of culture in evaluation is evident in the recent AEA Public Statement on Cultural Competent in Evaluation (April 2011), which emphasizes and re-affirms the importance of culture and cultural context in the evaluation field. This document highlights both the centrality and importance of cultural competence in evaluation theory and practice (AEA, 2011).[2] Furthermore, the presence of culture in evaluation can be found in the increasing literature on cultural context and culture in evaluation. For example, *The Role of Culture and Cultural Context: A Mandate for Inclusion, the Discovery of Truth and Understanding in Evaluative Theory and Practice* edited by Stafford Hood, Rodney Hopson, and Henry Frierson, feature a number of case studies in which evaluators employed approaches that were culturally responsive to, and respectful of, the cultural context of the programs that were

1 This is by no means an exhaustive list; however, there are a number of other evaluators of color who have contributed to the increased focus on culture in the evaluation field.

2 In April 2011, the AEA Board of Directors and the Task Force on Cultural Competence in Evaluation (TF-CC) asked members of AEA to vote on making the Public Statement on Cultural Competence in Evaluation public. The statement was developed as a result of a recommendation provided to the association by the Building Diversity Initiative to create a public education campaign that emphasizes the importance of culture context in evaluation.

evaluated. Additionally, several volumes of the *New Directions for Evaluation* journal have been dedicated to exploring issues of culture and context in evaluation, including special issues on co-constructing a contextually responsive evaluation framework (Volume 2004, Issue 101), cultural competence in evaluation (Volume 2004, Issue 102), and critical issues in STEM evaluation (Volume 2006, Issue 109).

Culturally Relevant Evaluation Approaches

As a result of the growing attention to culture in the evaluation field, a number of approaches that incorporate culture and context into the evaluation have emerged. Evaluation approaches that are responsive to the cultural context of the program in the evaluation process are known by a variety of names; however, there is a general agreement in the field that culture and context are important components in evaluations that attempt to be responsive to culture (Samuels & Ryan, 2011). Examples of approaches that incorporate the cultural context of the program in the evaluation include: culturally competent (SenGupta, et al., 2004), culturally responsive (Hood, et al., 2005), culturally consistent, transformative (Mertens, 1999; 2003), culturally sensitive, culturally anchored, and values-based (Greene, et al., 2006), multicultural, and cross-cultural (Chouinard & Cousins, 2009).

Evaluation approaches which respond to culture and context are rooted in participatory evaluation. Participatory evaluation approaches incorporate evaluation program stakeholders (i.e. administrators, staff, and participation) into the evaluation process (Cousins & Whitmore, 1998; King, 1998). According to Fitzpatrick, Sanders, and Worthen (2003) participatory evaluation approaches seek "first-hand experience with program activities and settings and involvement of program participants in evaluation" (p. 131). Therefore, evaluation approaches with a participant

orientation employ inductive reasoning. In addition, they use a multiplicity of data, do not follow a standard plan, and record multiple rather than single realities (p. 133-134).

For example, the Talent Development (TD) Evaluation Framework, an evaluation approach responsive to cultural context of the program, is rooted in the principles of participant evaluation. More specifically, the TD Evaluation Framework incorporates the major components of responsive evaluation, participatory evaluation, empowerment evaluation, and cultural component evaluation (Thomas & Stevens, 2004). Thomas and Stevens (2004) write, the TD Evaluation Framework is an evaluation approach that "is rooted in several traditions of evaluation that intentionally seek engagement with contexts of practice" (p. 1). The inclusion of stakeholders in the evaluation process from many participatory evaluation approach viewpoints, including culturally responsive evaluation, is an essential component of the evaluation process.

Culturally Responsive Evaluation (CRE)

Culturally responsive evaluation is an evaluation approach which seeks to be responsive to the cultural context of the program. Of the evaluation approaches which incorporate the program's cultural context into the evaluation, culturally responsive evaluation is one of the most widely known evaluation approaches (Samuels & Ryan, 2011). Culturally responsive evaluation values and appreciates the cultural context of the program. Additionally, it honors the culture of the participants in the evaluation. Evaluators who adopt a culturally responsive evaluation approach have a "responsibility to increase their awareness, knowledge, and appreciation of the culture of the program participants" (Askew, et al., 2012). Evaluators are also self-reflective in their practice and examine their own assumptions, biases and understandings in relation to the cultural context in which the evaluation takes place (Symonette, 2004).

Culturally responsive evaluation has been defined in different ways. As a theory, culturally responsive evaluation is "an extension of responsive evaluation that is markedly attentive to culture and context yet consonant with other evaluation approaches that attend to democracy, social change, inclusion issues, and power relationships" (Greene, 2006, as cited in Ryan, et al, 2011, p. 184). Further, the theory is "oriented around a set of values or ideology such as values-engaged evaluation (Greene, et al., 2005) and empowerment evaluation (Fetterman, 1994)" (Ryan et al., 2007, p. 201). In addition, culturally responsive evaluation is rooted in decolonizing frameworks for indigenous people (Hopson 2009; L.T. Simth, 1999), and race-based theories (Ladson-Billings & Tate, 1995) in which the evaluator is an advocate for social justice for disadvantaged and underserved populations (Askew et al., 2012).

As an evaluation approach, culturally responsive evaluation provides a set of guidelines for conducting evaluations (Askew, et al., 2012). For example, culturally responsive evaluation includes a "collection of practical skills and frameworks that attend to culture and context when preparing for an evaluation, conducting it, and disseminating and using the results of the study" (Samuels & Ryan, 2011, p. 185). Therefore, culturally responsive evaluation signals to the importance of a certain skills set that evaluators must have to be responsive to the cultural context of a program in the evaluation. This definition of culturally responsive evaluation is closely aligned with cultural competence.3 "At its most basic level, cultural competence reflects an awareness and appreciation of differences in cultural groups, and an ability to effectively communicate across cross cultural groups" (Botcheva, et al., 2009, p. 177). More precisely,

[3] Cultural competence and culturally responsive evaluation are two distinct evaluation approaches that oftentimes get mistaken for each other. However, it is important to note that although they share similar characteristics they are not the same; should be differentiated within the field.

> Cultural competence in evaluation can be broadly defined as a systematic, responsive inquiry that is actively cognizant, understanding, and appreciative of the culture in which the evaluation takes place; that frames and articulates the epistemology of the methodology; and that uses stakeholder-generated, interpretive means to arrive at the results and further use of findings. (SenGupta, et al., 2004, p. 13)

Similar to cultural competence, culturally responsive evaluation is an approach that focuses on the evaluator's ability to develop evaluations that are responsive to, and respectful of, cultural differences. From a culturally responsive evaluation perspective, an evaluator achieves cultural competence in the cultural context of the program when they have a full understanding of the "cultural norms, values, language and behavior codes" of the cultural groups in the program (Askew, et al., 2012). Further, cultural competency is required in culturally responsive evaluation. Evaluators cannot conduct an effective culturally responsive evaluation without a certain level of cultural competency in the culture of the program under evaluation (Frazier-Anderson, et al., 2012).

As a stance, culturally responsive evaluation seeks to equalize opportunities for individuals who have been underserved and/or undermined in evaluation efforts that have been driven by postpositivist paradigms (Parker, 2004). Further, evaluation efforts that have opted for methodological rigor over "a concern for the population studied" have limited the potential of evaluators to examine social context issues critical to the program (p. 88). Culturally responsive evaluation seeks to recognize the need for more evaluations with a social justice stance. Further, the purpose of the evaluation is to "give voice to persons whose perspectives are often ignored, minimized, or rejected" (Thomas & Stevens, 2004, p. 1). Thus, culturally responsive evaluation "requires an evaluator to raise issues of differential service delivery and accesses to race, gender, economic status, and power (Hood, 2009; Hopson, 2009)" (as cited in Askew, et al., 2012).

Key Tenets of Culturally Responsive Evaluation

While culturally responsive evaluation can be defined in different ways, there are three key tenents of culturally responsive evaluation: (a) culture, (b) context, and (c) responsive evaluation (Ryan, et al., 2007; Frierson, et al., 2002; Frierson, et al., 2010). The first tenet of culturally responsive evaluation is culture. In culturally responsive evaluation, evaluators incorporate the cultural context of the program in the evaluation. Further, the culture of the participants is considered an important factor in the evaluation (Frierson, et al., 2002; Frierson, et al., 2010). The second tenet of culturally responsive evaluation is context, an important component of the evaluation. The context of an evaluation can include the geographic location and social, economic, and political context in which the evaluation takes place (Thomas, 2004; Ryan, 2007). The third tenet of culturally responsive evaluation is responsive evaluation. In culturally responsive evaluation, evaluators are responsive to the program stakeholders. Further, "a major aspect of responsiveness is respecting, honoring, attending to, and representing stakeholders' perspectives" (House, 2001) (as cited in Thomas, 2004, p. 13). Below, a more thorough description is provided for each tenet of culturally responsive evaluation.

Culture. Culture can be defined in different ways. The most common definition in the literature includes the following definition.

> The word culture is defined as the shared values, traditions, norms, customs arts, history, folklore, and institutions of a group of people. In this case, culture shapes how people see their world and structure their community and other aspects of their lives. (Manswell-Butty, et al., 2004, p.39)

Culture can also be defined historically. As such, Ryan, et al., (2007) defined culture in the following way:

> As a historically transmitted pattern of meaning that has explicit and implicit expressions through symbols and beliefs and that is intertwined with an individuals' and groups' notion of identity across contexts (e.g. work place, school, race, and other identity-based groups. (p. 201)

Moreover, SenGupta et al., (2004) contended that culture is more than simple activities such as "as food, music, celebrations, holidays, dance, dress, and clothing" (p. 6). Rather culture is "rooted in inherent beliefs and values that influence customs, norms, practices, and social institutions, including psychological processes, language, caretaking practices, media, educational systems, and organizations" (p. 6). More comprehensively, Guzman (2003) highlighted that culture can take on many forms. It can be (a) an abstract human-made idea, (b) a context or setting in which individuals learn behaviors, (c) values, beliefs, attitudes, and languages specific to a group, and (d) passed on from generation to generation (p. 170). Further, Frazier-Anderson, et al., (2011) argued that culture is a difficult term to define; however, it is a "construct" that refers to an individual's "idea" and/or "belief" or an "activity," "place," or "thing" common among a group of people (p. 349).

While culture can be defined in different ways, there is a general agreement in the field that culture is "dynamic and fluid" (Frazier-Anderson, et al, 2011, p. 350; Kirkhart, 2010;). That is, culture is passed from one generation to next and can change depending on the external forces that present themselves in a particular context. In addition, culture is embedded within a specific set of historical, social, political, and economic contexts (Botcheva, et al., 2009). Lastly, culture can exist at multiple levels: individual and institutional level. Individual levels of culture refer to the ways in which individual's identify with "categorical labels attached to personal characteristics such as race, ethnicity, religion, and gender" (Kirthart, 2010, p. 401). Culture at the institutional level refers to the ways in which institutions impart "knowledge, values, beliefs, and skills that are also shared and communicated across cohorts or generations" (p. 401).

Context. Context refers to factors present in the evaluation context that can impact the design, implementation, or evaluation of a program. According to Thomas (2004), context

"includes the combination of factors (including culture) accompanying the implementation and evaluation of a project that might influence its results" (p. 11). The context of the evaluation can include geographical location, timing, political and social climate, and economic conditions (Thomas, 2004). Context can also comprise the site, location, and environment of the program. Additionally, context can include the: (a) material conditions of the setting; (b) institutional and organizational climate of the funding agency; (c) personal and interpersonal relationships between the participants; and (d) political dynamics of the program such as power, privilege, and influence (Greene, 2005). Therefore, context can refer to multiple levels of interaction between individuals and contexts (i.e. federal, state, and local).

For some scholars in the field, "context is undervalued, often taking a secondary rather than primary role, particularly in effectiveness evaluation" (Rog, 2009 as cited in Kirkhart, 2010, p. 405). Strategies for attending to, and better understanding the cultural aspects of the context of a program are outlined below.

1. Learn the history of this community and of the evaluand within in.
2. Identify the relevant geographic boundaries and characteristics of this context.
3. Reflect on shared aspirations, values, and ideals and how they are represented in the institutions and governance structures of this context.
4. Identify the strengths of this context. Notice conditions or circumstances are identified as problematic in this context.
5. Notice how power is distributed in this context through both formal and informal structures.
6. Notice what cultural characteristics are the most salient in understanding the consumers of this program and the communities from which they come.

7. Notice what cultural characteristics are most salient in understanding the providers of this program and the communities from which they serve.
8. Consider the values espoused by the funders of this evaluand.
9. If this program is built on prior empirical research, notice who participated in the original studies that comprise this body of evidence and examine how culture was addressed in that earlier work. (Kirthart, 2010, p. 405)

Responsive evaluation. Responsive evaluation is an evaluation approach that addresses key "issues" relevant to a particular evaluation context (Fitzpatrick, Sanders, & Worthen, 2005). Responsive evaluation incorporates three main components. First, responsive evaluation attends to program activities more than program intents. Two, responsive evaluation responds to audience requirements for evaluation information. Further, responsive evaluation seeks to be useful to the program stakeholders, and therefore incorporates program stakeholders in the evaluation process. From the perspective of responsive evaluation, program stakeholders can include the individuals who have either a personal or financial stake in the evaluation results. Three, responsive evaluation incorporates the different value-perspectives held by stakeholders in the evaluation and value the perspectives of program stakeholders in the evaluation. According to Stake, "in these three separate ways an evaluation plan can be responsive" (Stake, 1975, p. 5).

In addition to the three components above, responsive evaluations are developed based on issues of importance to the program stakeholders. Additionally, the purpose of a responsive evaluation is shaped by issues that arise within the evaluation context. Further, stakeholder participation is a key component of responsive evaluation: "From a responsive evaluation perspective, stakeholders' involvement in the program evaluation is important because these

individuals have a wealth of knowledge to share" (Manswell-Butty, et al., 2004, p. 39). For example, program stakeholders have knowledge on the program, participants and program context. In addition, they are familiar with the clients they serve, and are attuned to the participant interests in the program (Askew et al., 2012). Therefore, in responsive evaluation, it is believed that "those who carry out the program, who are participating in it, have much more knowledge than the evaluator who still has to discover what is going on" (Abma & Stake, 2001, p. 15).

Stake (2004) describes responsive evaluation as an approach to evaluation in which is attuned to the program context. That is, responsive evaluation pays close attention to the types of activities that take place in the evaluation context (i.e., interactions between the program staff and participants). Responsive evaluation provides the "experience of personally being there, feeling the activity, the tension, knowing the people, and their values" (p. 86). Therefore, responsive evaluation is particularly attentive to what occurs within the context of the program: the activities or the interactions of the participants. Further, responsive evaluation seeks to be "in touch with multiple but particular positions, interest and perspectives throughout the evaluation process" (Samuels & Ryan, 2011, p. 85).

The primary research method in responsive evaluation is the case study. Case study research determines program quality through precise documentation of the program context. Furthermore, data on program quality is collected via methods of precise documentation: a) observations, b) focus groups, c) interviews, and d) other qualitative methods. In case study research, decisions on program quality are often made on the basis of data collected over a significant amount of time. More specifically, observations are one way to provide stakeholders with information about key activities (Abma & Stake, 2001).

There are two types of case studies: a) instrumental and b) intrinsic (Stake, 1995). The goal of an instrumental case study is to generate knowledge and understanding of a particular issue(s) or problem(s). Therefore, instrumental case studies investigate one particular case in order to inform a larger phenomenon of interest. The goal of an intrinsic case study is to understand a phenomenon in one particular case. Intrinsic case studies examine the uniqueness of the phenomena of interest in the case study. Both case study methods examine the context of the case or issue in the study. Further, cultural, organizational, curricular, legal, professional, policy, and collegial factors are taken into account (Abma & Stake, 2001). Additionally, other factors such as the experiences of the individuals with the program and perceptions of the program stakeholders are examined in order to document the complexity of the issue or case in the study.

There are many ways culturally responsive evaluation builds upon the tenets of responsive evaluation. First, culturally responsive evaluation includes program stakeholders in the evaluation process. Therefore, culturally responsive evaluation "is intended to extend Stake's idea of 'responsiveness' to include all evaluation stakeholders via strategies of inclusion" (Ryan, et., 2007, p. 202). Program stakeholders are treated as "team members, designers, and implementers of the evaluation" (p. 202). Second, culturally responsive evaluation responds to the complex nature of the program. Culturally responsive evaluation "recognizes that within the evaluation context, there are different dimensions, locations, and perspectives, and characteristics of culture that influence the ways programs are designed, implemented, and experienced by individuals and groups" (Samuels & Ryan, 2011, p. 185). Thirdly, culturally responsive evaluation recognizes the importance of a shared lived experience and honors the cultural context of the evaluation by employing evaluation strategies aligned with the experiences of the participants (Askew, 2012; Frierson, et al., 2002; Frierson, 2010).

Shared lived experiences. Shared lived experiences are an important component of culturally responsive evaluation. From this evaluation perspective, shared lived experiences "provide a sense of what it means to live and exist through the eyes of another" (Askew, et al., 2012, p. 3). Shared lived experiences refer to "a collective understanding of cultural norms, values, language, and behavioral codes" (p. 3), and they provide certain benefits for the evaluation. For example, evaluators who have shared lived experiences are more likely to understand the program context. More specifically, evaluators are more attuned to the racial, social, and economic conditions dynamics of the program. Additionally, shared lived experiences provide evaluators with a greater ability to "truly hear" what participants say and mean. For example, "knowing the language of a group's culture guides one's attention to the nuances of how language is expressed and the meaning it may hold beyond mere world" (Frierson, et al., 2002, p. 71). Further, African American evaluators are more likely to be attuned to the non-verbal behaviors of African American students; the same is true for a Latino/an evaluator in a Latino evaluation context.[4]

A shared lived experience is particularly important in the data analysis and interpretation phases of an evaluation. For example, evaluators who have a shared lived experience are more likely to analyze and interpret the data accurately. Evaluators with a shared lived experience are also more likely to determine accurate meaning(s) of the data. As such, "determining meaning of what has been observed is central in culturally responsive evaluation" (Frierson, et al., 2002, p. 70). Furthermore, possessing an understanding of the cultural context of the program is also necessary in making accurate judgments about program quality. At times, possessing an adequate understanding of the cultural context is not enough; rather a sensitivity of the cultural context of

[4] Implications for understanding cultural nuances can be found in other fields outside of evaluation. For example, the fields of teaching, medicine, communication and business/marketing all stress the importance of attending to culture (Pewewardy, 1994; Philips; 1983; Hartwig, 2010; Page, 2005; Hilton, 2007).

the program is also required in culturally responsive evaluation. As such, culturally responsive evaluations are designed to "honor the cultural context in which the program takes place by bringing needed shared lived-experiences and understanding to the evaluation" (Ryan, et al., 2007, p. 201).

When evaluators do not have a shared lived experience, the evaluation is at risk (Frierson et al., 2002; Frierson, 2010). For example, evaluators who do not possess an understanding of the cultural context of a program are more likely to misinterpret the evaluation data and ultimately make unsound judgments about program quality. Moreover, evaluators who do not possess some level of familiarity with a group's behaviors, cultural norms, and traditions are less likely to analyze and interpret the evaluation accurately. Thus, culturally responsive evaluation seeks to increase the number of evaluators from minority backgrounds in order to ensure that the evaluation field adequately addresses the needs of a diverse and growing population in the United States. As the United States becomes more and more diverse, the evaluation field will also be challenged to develop and adopt new approaches for evaluating programs. Thus, culturally responsive evaluation is one evaluation approach that meets this challenge.

Roots of Culturally Responsive Evaluation

In many ways, culturally responsive evaluation is a participant-oriented evaluation approach and seeks stakeholder participation in the evaluation. More specifically, culturally responsive evaluation builds on Robert Stake's (1995) notion of responsive evaluation. Culturally responsive evaluation extends the notion of responsive evaluation to include cultural context in evaluation. Hood and Hopson (2008) purport:

> Stake's articulation of responsive evaluation (1995) has been at the core of the work that we have undertaken to expand the concept of that of culturally responsive evaluation. The intent is to extend the boundaries of responsive evaluation to include the tradition of

> African American evaluative inquiry, research, and practice in the field of evaluation. (p. 411)

Culturally responsive evaluation is most closely aligned responsive evaluation; however, is also rooted in several other evaluation traditions: a) deliberative democratic evaluation (House & Howe, 1999, House, 1999, House & Howe, 2000; House, 2001; House & Howe, 2001), b) participatory evaluation (Cousins & Whitmore, 1998), c) empowerment evaluation (Fetterman, 1994), d) inclusive evaluation (Mertens, 1999, 2003; Mertens & Hopson, 2006), and; f) multicultural validity (Conner, 2004; Kirkhart, 2005; Nelson-Barber, LaFrance, Trumbell, & Aburto, 2005; Kirkhart, 2010). To illustrate how culturally responsive evaluation is rooted in these evaluation traditions, a description for each of these traditions is provided below.

Deliberative democratic evaluation. Deliberative democratic evaluation is a participatory evaluation approach in which seeks to advance principles of democracy in evaluation (House & Howe, 1999, House, 1999, House & Howe, 2000; House, 2001; House & Howe, 2001). Democratic evaluation rests on the principles of democratic theory, which assumes that evaluation should be participatory in nature: "The deliberative democratic approach to evaluation is grounded in deliberative democratic theory, which adopts a relatively strong stance toward stakeholder participation" (Howe & Ashcraft, 2005, p. 2275). Furthermore, approach advocates for genuine democracy, which includes equal representation for all key evaluation stakeholders in the evaluation process. Deliberative democratic evaluation is particularly interested in providing voice to key groups who have traditionally been marginalized in evaluation: "Therefore, those program stakeholders whose voices have typically been minimized in the evaluation process – as a result of their being the least powerful in terms of race, gender, and/or socioeconomic status – are given equal voice in the process" (Hood & Hopson, 2008, 415).

The three principles of deliberative democratic evaluation are: (a) *inclusion*, (b) *dialogical*, and (c) *deliberation* (House & Howe, 1999; House, 1999; House & Howe, 2000). First, evaluators should design evaluations that are inclusive of all relevant interests. Evaluators should not design evaluations that are only representative of one particular group (i.e. the most powerful group) (House & Howe, 2000). In democratic societies "all those who have legitimate, relevant interests should be included in decisions that affect those interests" (p. 5). Thus, in deliberative democratic evaluation, the relevant interests of all stakeholder groups are not only considered but also represented in the evaluation.

A major challenge evaluator's face in designing deliberative democratic evaluations is power imbalances that exist within evaluation contexts. Power imbalances may impede an evaluator's ability to represent all relevant interests. For example, "the powerful may dominate the discussion, or those without power may not be represented adequately" (House & Howe, 2000, p. 6). Thus, it is an evaluator's responsibility to address these power imbalances by providing voice to all those who may not be able to do so themselves due to lack of power:

> Evaluators must design evaluations so that relevant interests are represented and so that there is some balance of power among them, which often means representing the interests of those who might be excluded in the discussion, because their interests are likely to be overlooked in their absence. (House & Howe, 1999, p. 98)

Moreover, Hood (2000) acknowledges that true representation of all relevant interests may require that evaluators acknowledge that "the realities of democracy in the United States have been experientially different for its oppressed groups" (p. 77). Thus, evaluators, although well intentioned, may have a difficult time implementing principles of social justice in evaluation because they are unfamiliar with, or unaware of, how groups who have traditionally been underserved in evaluation experience democracy. Moreover, Hood (2005) calls for the representation of more evaluators of color who may be "instrumental in determining and

weighting stakeholders' interests" (p. 88). Hood highlights the importance of common experience in these evaluations:

> The ability to look behind statements of need to the values that guide them, in the case of racial minorities in American, is more likely to occur if the evaluation can draw on shared lived experiences as a result of cultural background or extended experiences within a particularly culture. (p. 81)

Second, evaluators should design evaluations that are dialogical in nature. That is, evaluations should be structured to provide opportunities for individuals and groups to engage in dialogue. Through dialogue, evaluators have the critical task of determining the real interests of a particular group or individual. Real interests are different than perceived interests: "It may be that through dialogue and deliberation stakeholders will change their minds as to what their real interests are" (House & Howe, 2000, p. 7). Through dialogue, the deliberative democratic evaluator can learn about the "position, views, and interests" of various stakeholders (p. 7). Dialogue is difficult to achieve in some environments, for example in places where programs and policies are complex in nature. Therefore, "The more complex the situation, the more dialogue is needed to sort it out" (p. 7).

Third, democratic evaluations should be deliberative. That is, deliberation "is fundamentally a cognitive process, grounded in reasons, evidence, and principles of valid argument" (House & Howe, 2000, p. 8). Deliberation is pursued to ensure that the real interests and values of the stakeholders are revealed. The role of the evaluator in the deliberation phase of the evaluation is to weave out values, preferences, and interests that are factual (as opposed to perceived) and to maximize the factual as best they can (Howe & Howe, 1999). As such, evaluators should not assume the values or interests of the evaluation stakeholders. Rather, values should be revealed via deliberation or valid argument. Furthermore, evaluations should

not replace traditional decision-making procedures (i.e. voting); however, they should be used to inform decisions made from the evaluation results (p. 8).

Participatory evaluation. Participatory evaluation is a collaborative evaluation approach which seeks participation with individuals and/or groups in the evaluation. There are many ways in which participatory evaluation has been defined in the literature. As such, the term:

> is used quite differently by different people. For some it implies a practical approach to broadening decision making and problem solving through systematic inquiry; for others, reallocating power in the production of knowledge and promoting social change are the root issues. (Cousins & Whitmore, 1998, p. 5)

While there are many ways that participatory evaluation is defined, yet it is typically characterized by the inclusion of stakeholders in the evaluation. Stakeholders are individuals or a group of individuals who have a vested interest in the program or policy being evaluated. Thus, stakeholders can include program sponsors, managers, developers, implementers, members of special-interest groups, and problem beneficiaries (Cousins & Whitmore, 1998).

Cousins and Whitmore (1998) posit that there are two "streams" of participatory evaluation: a) practical participatory evaluation (P-PE) and b) transformative participatory evaluation (T-PE). First, the goal of P-PE is to engage stakeholders in the evaluation. Under P-PE, stakeholder participation in the evaluation increases the evaluation's "relevance, ownership, and thus utilization" (p. 6). Stakeholder participation can have three types of impact on the evaluation context: (a) instrumental, (b) conceptual, and (c) symbolic. For example, instrumental impact would help in decision-making processes. Two, conceptual impact would be educative and serve a learning function in the evaluation context. Three, symbolic impact would advance a particular decision or agenda. Under certain circumstances, P-PE not only enhances the utilization of a particular set of evaluation results or evaluation, but it also enhances the impact of the evaluation in the evaluation context. Furthermore, "it has been demonstrated that under

appropriate conditions participation by stakeholders can enhance utilization without comprising technical quality or credibility" (p. 7).

Transformative participatory evaluation (T-PE) promotes a social justice agenda by addressing "participatory principles and actions in order to democratize social change" (Cousins & Whitmore, 1998, p. 7). The goal of T-PE is to empower individuals who have been either ignored or under-represented by traditional evaluation frameworks. Brisolara (1998) defines transformative-evaluation as an approach to evaluation that "responds to evaluation's interests in the political nature of evaluation" (p. 32). In T-PE, the role of the evaluator is to promote a social justice agenda and advocate for populations who have traditionally had less power in the evaluation process. Furthermore, the notion of objectivity is overlooked by TP-E supporters who believe that "objectivity itself is a product of a colonizing science" (p. 34).

There are three main components of TP-E. First is the notion of knowledge process: who produces knowledge and whose knowledge is valid in the evaluation? Second is the evaluation process: who plays a role in the evaluation process, who does not play a role and why?. Third is critical reflection, which "requires participants to question, to doubt, and to consider a broad range of social factors, including their own biases and assumptions" (Cousins & Whitmore, 1998, p. 8).

Empowerment evaluation. Empowerment evaluation seeks to empower individuals to evaluate programs within their own organizations. Empowerment evaluation "focuses on improvement, is collaborative, and requires both qualitative and quantitative methodologies" (Fetterman, 1994, p. 1). Empowerment evaluation is done for various reasons: (a) implementation, (b) improvement, and (c) assessment. However, the primary purpose is improvement. Fetterman (1994) emphasizes empowerment at the individual level: "The focus is

on helping people help themselves" (p. 1). Thus, empowerment evaluation seeks to provide individuals with an array of tools, skills, and strategies for improving their own programs.

Fetterman (1994) defines empowerment evaluation as "the use of evaluation concepts and techniques to foster self-determination" (p. 1). Stakeholder participation is an important component of empowerment evaluation. Empowerment evaluation is collaborative in nature, engaging stakeholders in the evaluation process. This approach is multidisciplinary and can be used in variety of disciplines, such as health, education, business, agriculture, microcomputers, non-profits and foundations, government, and technology. More specifically, empowerment evaluation employs various methodologies, qualitative and quantitative (Fetterman, 1994). Under an empowerment evaluation approach, different methodologies provide a more comprehensive picture of the phenomena under study, as opposed to evaluation approaches that employ only one method of data collection.

The major tenants of empowerment evaluation include: a) training, b) facilitation, c) advocacy, d) illumination, and e) liberation (Fetterman, 1994). Firstly, empowerment evaluation can provide training to individuals or organizations to conduct their own evaluation: "This approach desensitizes and demystifies evaluation and ideally helps organizations internalize evaluation principles and practices, making evaluation an integral part of program helping" (p. 3). Secondly, empowerment evaluation can serve to facilitate the evaluation process for individuals and organizations alike. As such, "evaluators can serve as coaches or facilitators to help others conduct their evaluation" (p. 4). Thirdly, it can provide advocacy to individuals and programs. Evaluators can serve as "direct advocates" for groups or programs (p. 6). Fourthly, this type of evaluation can be illuminating, providing stakeholders with new and interesting information about a particular program or policy. Fifthly, empowerment evaluation can help in

the liberation of individuals or organizations. For example, empowerment can help "individuals take charge of their lives – and find useful ways to evaluate them-liberates them from traditional expectations and roles" (p. 9).

Inclusive evaluation. Inclusive evaluation seeks to include marginalized groups in the evaluation process. Inclusive evaluation is based on transformative theory:

> Transformative theory is used as an umbrella term that encompasses paradigmatic perspectives such as emancipatory (Lather, 1992; Mertens, 1998) antidiscriminatory (Humphries & Truman, 1994; Truman, Mertens, & Humphries, 2000), participatory (De Koning & Marion, 1996; Reason, 1994; Whitmore, 1998, and Freirian approaches (McLaren & Lankshear, 1994). (Mertens, 2003, p. 93)

Transformative theory works in direct contrast to traditional evaluation paradigms. Transformative theory believes that knowledge is not neutral, but rather a reflection of unequal power and social relationships in society (Mertens, 1999). Thus, inclusive evaluation takes on a participatory evaluation approach in which the interests, concerns, and needs of marginalized populations are an important component of the evaluation. Marginalized groups can include individuals or groups of people who have in one way another experienced discrimination based on gender, race, or socio/economic status (Mertens, 1999).

There are both strengths and limitations to inclusive evaluation. First, inclusive evaluation can help in "aiding impartiality, gaining new information, and improving the sensitivity of evaluation" (Mertens, 1999, p. 3). Additionally, inclusive evaluation meets the evaluation guidelines set forth by various professional organizations (i.e. AEA) for evaluators to be both moral and ethical in their work. Moreover, inclusive evaluation allows individuals to meet their symbiotic obligation to other individuals in society (p. 3). Lastly, inclusive evaluation gives an accurate representation of the program and its activities because the evaluation includes

the perspectives of more than one group of individuals. Because it is inclusive on many level, inclusive evaluation seeks to provide more accurate evaluation results.

There are various limitations to inclusive evaluation. First, evaluators may experience challenges in convincing stakeholders to participate in the evaluation. Evaluation stakeholders' participation is not easy to achieve in every evaluation and will largely depend on the circumstances of the evaluation. In some cases, stakeholders' participation may prove impossible to achieve in the evaluation. Moreover, stakeholders' participation in the evaluation must be meaningful. Meaningful engagement in some evaluations may also be difficult to achieve for various reasons. Additionally, evaluators may encounter challenges in maintaining a balanced role in the evaluation, as an evaluator and participant. In inclusive evaluation, the evaluator must remain responsible to the evaluation stakeholders and ensure that their biases do not interfere with their work in the evaluation (Greene, Lincoln, Mathison, Mertens, & Ryan, 1998).

Multicultural Validity. Multicultural validity has been addressed in conjunction with other types of validity in the social sciences. Validity is the ability to make accurate conclusions or inferences about a given population. In evaluation research, validity can be used to determine to what extent the evaluation results are valid. There are several different types of validity: a) conclusion validity, b) internal validity, c) construct validity, and 4) external validity. In social science research, validity is important because it allows researchers to apply findings from the target population to other populations in similar contexts.

Multicultural validity represents the intersection between evaluation theory and cultural context. In evaluation, multicultural validity is used to determine to what extent the evaluation theory used is congruent with, and accurately represents, the cultural context under study. According to Kirkhart (2010), "multicultural validity refers to the accuracy or trustworthiness of

understandings and judgments, actions, and consequences, across multiple, intersecting dimensions of cultural diversity" (p. 401). Furthermore, Kirkhart (2010) argues that multicultural validity is not a new concept in social science research but encompasses similar dimensions of other types of validity; the author adds the dimension of cultural diversity in evaluation. Multicultural validity assess "how well evaluation captures meanings across dimensions of cultural diversity, and it scrutinizes the accuracy or trustworthiness of the ensuing judgments of merit and worth" (Kirkhart, 2005, p. 22) Thus, multicultural validity requires an evaluator to employ an evaluation theory that is congruent with the cultural context of the program. Employing a cultural theory that is congruent with the evaluation context "presumes that one understands both the cultural location of theory and the cultural dimensions of context" (Kirkhart, 2010, p. 401).

In addition to multicultural validity, issues related to reliability have also been addressed in the context of culturally responsive evaluation. For example, Nelson-Barber, et al., (2005) examined validity and reliability within evaluation theory and practice. In examining various examples of evaluations that addressed cultural issues, the authors called for the inclusion of cultural competence in both evaluation theory and practice. More specifically, they highlighted the need for evaluators to be knowledgeable of diverse perspectives, values, and lifestyles across various ethnic groups. To obtain results that are both reliable and valid, evaluators must be attuned to cultural factors related to different programs. When evaluators lack a complete understanding of the cultural context of a program, they are more likely to "miss important information that can shed light on why a program has particular outcomes or impact on a community" (Nelson-Barber, et al., 2005, p. 62).

Reponses to Culturally Responsive Evaluation

There have been both positive and negative responses to culturally responsive evaluation in the evaluation field. Evaluators who support culturally responsive evaluation believe that it is not only necessary, but also an important component of program evaluation in the 21st century. There are some scholars who have challenged the basic premises of culturally responsive evaluation. More specifically, there are two main arguments against culturally responsive evaluation. The first argument is that evaluations should not take culture into account and evaluations should be culture-free. The second argument is that evaluators should not be responsive to cultural context. For example, "some individuals argue that while an evaluation should take into account the culture and values of the project or program it is examining, it should not, however be responsive to them" (Frierson, et al.,, 2002, p. 64).

Both arguments above have been addressed by scholars in the evaluation field. Precisely, both arguments are refuted in the NSF 2002 and 2010 User-Friendly Program Evaluation Manuals. In relation to the first statement, Frierson, et al., (2002) imply that "just as surely as there no culture-free evaluators, there are no culture-free evaluations, educational tests, or societal tests" (p. 63). Therefore, the question that evaluators should ask themselves is: whose values should be promoted in an evaluation? Frierson, et al., (2002) further state that values are part of everyday life and evaluation practice as well; thus our values as evaluators are reflected in our everyday actions and behaviors. As evaluators, we promote a certain set of values in our work; culturally responsive evaluators also "bring their cultural experiences to the evaluation; thus their evaluations reflect their own work" (p. 185). Because values are embedded in our work, it is important for evaluators, regardless of which evaluation paradigm or framework they employ, to acknowledge that their own personal cultural biases and assumptions are reflected within their practice. Furthermore, culturally responsive evaluation holds educational evaluators

accountable for "their own personal cultural preferences and to make a conscious effort to restrict any undue influence they might have on the work" (Frierson, et al., 2002. p. 64).

As Frierson, et al., (2002) imply, a culturally responsive evaluation "is based on an examination of impacts through lenses in which the culture of the participants is considered an important factor, thus rejecting the notion that assessments must be objective and culture free, if they are to be unbiased" (p. 63). As such, all evaluation is subjective in one form or another because our values as evaluators are reflected in our work. Recognizing that there are no culture-free evaluations, the purpose of culturally responsive evaluation is not to provide an "objective" evaluation; it is to accurately assess program quality by taking into account the cultural context of the program in an evaluation. When applied, culturally responsive evaluation has been shown to have a positive impact on the project and stakeholders (Manswell-Butty, et al., 2004). Furthermore, to undermine the notion that the cultural context of a program impacts the evaluation "is to put the program in danger of being ineffective and to put the evaluation in danger of being seriously flawed" (Frierson, et al., 2002, p. 63).

Frierson, et al., (2002) address the second argument by stating that while evaluators can provide knowledge as to how a program operates and whether or not it is successful, it remains necessary for evaluators to also understand the cultural environments in which the evaluation takes place (i.e. social, economic, and political circumstances). To not recognize how social, political, and economic factors impact the cultural context of a program is to provide a disservice to the program and its participants. Furthermore, while an evaluation can consider the cultural context of a program, culturally responsive evaluation advocates for more than consideration, but actual responsiveness to the cultural context of a program. Finally, Frierson, et al., (2002) argue, "it is one thing to accept or recognize the reasonableness of the requirement to describe the

cultural context. It is quite another to adopt evaluation strategies that are consonant with the cultural context(s) under examination" (p. 64).

Overall, culturally responsive evaluation rejects the arguments above on the basis that all evaluations are shaped by an evaluator's cultural biases and assumptions. As mentioned above, culturally responsive evaluation rejects the notion that evaluation should be culture-free and objective. Thus, culturally responsive evaluation not only accounts for but also responds to the cultural context of the program. The role of values in evaluation has been a highly debated topic. According to Greene, et al., (2006) "evaluators have always attended to values in evaluation, because it is fundamentally assessment of quality or 'goodness,' though more often descriptive than prescriptive'" (p. 59). Thus, values are always a part of an evaluation and therefore are a part of the fabric of the cultural context of the program. Culturally responsive evaluation recognizes that "inherent beliefs and values orientations" are rooted in culture and therefore seek to address the cultural context of the program. Furthermore, SenGupta, Hopson, & Thompson-Robinson (2004) suggest that "a common thread between culture and evaluation is the concept of values" (p. 6).

Culturally Responsive Evaluation Practice

This dissertation seeks to explore how evaluators plan, design, and implement culturally responsive evaluation. As such, it focuses on the practical dimensions of this type of evaluation. Practically, culturally responsive evaluation can be defined as "a collection of practical strategies and frameworks that attend to culture and context when preparing for an evaluation, conducting it, and disseminating and using the results of the study" (Samuels & Ryan, 2011, p. 185). The following section examines one evaluation framework that can be used to develop culturally responsive evaluations, the Talent Development (TD) Evaluation Framework (Thomas, 2004). In

addition, it reviews guidelines and strategies for conducting culturally responsive evaluations as included in the National Science Foundation's 2002 and revised 2010 User-Friendly Handbook for Project Evaluation Manual (Frierson, et al., 2002; Frierson, 2010).

The talent development evaluation framework. The Talent Development (TD) evaluation framework, developed by the Center for Research on the Education of Students Placed at Risk (CRESPAR), provides evaluators with an evaluation framework for designing and implementing culturally responsive evaluations. CRESPAR seeks to assist evaluators with the development and implementation of evaluation projects that are culturally and contextually responsive. Based on the Talent Development (TD) School Reform model, the TD evaluation framework incorporates the major tenets of participatory evaluation approaches including: (a) responsive, (b) participatory, (c) empowerment, and (d) culturally competent. The TD School Reform Model and the TD evaluation framework are reviewed in this chapter because they can be useful in developing and/or implementing culturally responsive evaluation.

The TD evaluation framework "is grounded in well-accepted evaluation concepts and principles" (Thomas, 2004, p. 1). Its primary goal is to provide a practical template for evaluators who evaluate programs and interventions in urban contexts. Additionally, it seeks to empower individuals in TD interventions. Thus, the TD evaluation framework endeavors to be responsive to the participants of the program or intervention. More specifically, the TD evaluation framework focuses on low income students in urban school contexts. As such, the approach endeavors to give voice and representation to underrepresented student populations who have been "ignored, minimized, or rejected in urban school settings" (p. 1).

The TD Model of School Reform has six major themes (Thomas, 2004, p.3). The themes reflect CRESPAR's focus on students in urban school settings. It aims to provide students with a "vast

array of intellectual, social-emotional; and transformative competencies; character building; personal fulfillment; and dual competence in the larger society as well as in their local community" (p. 5).

1. Builds on students' assets;
2. Provides students with transitional support across key developmental periods in their lives;
3. Engages students in constructivist and activist learning;
4. Prepares students with skills for careers for the twenty-first century;
5. Promotes the concept of school as community; and
6. Focuses on the meaning and connected learning.

The Talent Development (TD) evaluation framework builds on the tenants of the TD School Reform Model. The TD evaluation framework also includes the major principles of participatory evaluation, as well as some "methodological, conceptual, and strategic" issues are addressed in this evaluation framework (Thomas, 2004, p. 7). The following five components are included in TD evaluations; are described in the below in further detail.

1. Engaging stakeholders;
2. Co-construction;
3. Culturally and contextual relevance;
4. Responsiveness; and
5. Triangulation of conceptual and methodological perspectives

Descriptions for each of the components in the TD evaluation framework are provided below.

Engaging stakeholders. Like other participatory-oriented evaluation approaches, TD evaluations seek to engage stakeholders in the evaluation process. Engaging stakeholders is an

essential component of TD evaluations. Stakeholder participation is not always easy to achieve in an evaluation, particularly in evaluation contexts where there are underrepresented populations. There is, however, a variety of strategies evaluators can employ to achieve stakeholder participation in an evaluation. Evaluators can build trustworthy relationships with stakeholders, and they can be respectful of the various cultures present of the evaluation. Lastly, evaluators can develop partnerships with key stakeholders.

Co-construction. Co-construction directly relates to engaging stakeholders in the evaluation process. Co-construction refers to an evaluators' ability to collaborate with key stakeholders in the conceptualization, implementation, and evaluation of a program (Thomas, 2004). To establish co-construction, evaluators develop relationships with key stakeholder groups within the community, such as educators, school administrators, students, and families. Evaluators also establish positive relationships with TD project designers and implementers in order to ensure that they too are included in decisions about the evaluation.

Co-construction emphasizes the importance of power relations between and among key stakeholder groups in the evaluation. To ensure that all stakeholders are equally represented, co-construction seeks to equalize power relations among powerful and less powerful groups. Thomas (2004) expresses: "co-construction by necessity, involves a redistribution of power, assuming a kind of equality among different stakeholders" (p. 9). For example, co-construction provides equal opportunities for each stakeholder group to be represented within each phase of the evaluation. In TD evaluations, every individual's opinion is respected, valued, and appreciated.

The notion of co-construction is based on the idea that inquirer-stakeholder collaboration is an essential component of culturally responsive evaluation. Authentic collaborative

relationships between evaluators and stakeholders aid in the development of trust and respect across all stakeholders. In addition, "co-construction" takes into account social and cultural issues not addressed by other participatory approaches of evaluation (Cousins & Whitmore, 1998). Co-construction provides a framework in which evaluators can engage stakeholders in decisions regarding the evaluation questions, the evaluation purpose(s), data collection methods, and the dissemination of the results (Thomas, 2004).

Culture and contextual relevance. The TD evaluation framework aims to be culturally relevant and responsive to the cultural context of the program. The cultural context of a program is taken into account in the evaluation. In TD evaluations, culture refers to "shared values, traditions, norms, customs, arts, history, folklore, and institutions of a group of people" (Thomas, 2004, p. 11). The TD evaluation framework emphasizes the importance of being culturally competent. Culture competency refers to one's ability to relate, understand, and appreciate other cultures (SenGupta, et al., 2004; Thomas, 2004). In addition, the TD evaluation framework examines the context in which the program operates. Context refers to the "geographical location, timing, political and social climate, and economic conditions" (Thomas, 2004, p. 11).

Responsiveness: A central component of culturally responsive evaluation is the inclusion of relevant stakeholders in all phases of the evaluation process. Thomas (2004) states, "key stakeholders must be engaged in authentic ways throughout the entire process" in order to ensure that all key stakeholders voices are represented in the evaluation (p.5). In the TD evaluation framework, the voices of underrepresented groups are of particular importance. Culturally responsive evaluations include stakeholders in decisions regarding the evaluation design, questions, instrumentation, data collection, and dissemination of the findings.

Triangulation of perspectives: Triangulation is not a new phenomenon in social science research (Thomas, 2004). Triangulation is especially important with the use of mixed methods. Greene (2007) and other mixed methodologists have studied the benefits of triangulation in mixed methods and other studies. Studies have found that triangulation, when used properly, can increase the validity and reliability of the results (Greene, 2007).

In the field of mixed methods, triangulation includes the employment of more than one method to tap into the same facets or dimensions of the construct that is being investigated (Greene, 2007). Similarly, Thomas (2004) writes: "Triangulation generally involves the use of both quantitative and qualitative data in a single study or in multiple studies of a sustained program of research on a particular phenomenon" (p. 14). Triangulation calls for the use of multiple methods, both qualitative and quantitative, to study a particular phenomenon. For example, in a study investigating students' perceptions toward school, a researcher could conduct qualitative interviews with students as well as distribute student surveys with close-ended questions that access student perceptions towards school. By measuring the same phenomena via multiple methods (more than one, qualitative and quantitative), triangulation increases both the validity and reliability of a study's results.

The TD evaluation framework employs multiple types of triangulation. They include: a) investigator triangulation, b) multiple operationalism, c) methodological triangulation, d) target person triangulation, and; e) analysis triangulation (Thomas, 2004).[5] By using multiple types of triangulation, TD evaluations access the same phenomenon in urban schools from multiple

[5] A complete description of each triangulation strategy is included in: Thomas, V. (2004). Building a Contextually Responsive Evaluation Framework: Lessons from Working with Urban School Interventions, *New Directions for Evaluation, 101*, 3-19.

perspectives. For example, target person triangulation examines issues from multiple perspectives – teachers, parents, and program managers – in order to achieve a more comprehensive outlook of the context and culture of a program or school. Moreover, investigator triangulation includes the employment of more than one researcher with similar interests, researchers with different areas of expertise. In TD evaluations, triangulation is used to increase the "strengths of interpretations and conclusions drawn" from TD evaluation results (Thomas, 2004, p. 15). Moreover, triangulation is not sought for the purpose of obtaining "a single truth," as is common other research. Rather, in TD evaluations triangulation is employed to achieve "inclusiveness of perspectives and validity through this inclusiveness" (Thomas, 2004, p. 15).

In sum, the TD Evaluation Framework provides an evaluation framework for conducting evaluations that are both culturally and contextually relevant to the program's context. This framework can be applied to conduct evaluations in various fields (i.e., social work, health, and community health); however, this framework is targeted at evaluating intervention efforts in urban school districts. According to Thomas (2004), this type of evaluation approach is necessary when evaluating programs in traditionally underserved communities. Perhaps the greatest strength of this evaluation framework is its commitment to serving underrepresented populations in programs. As Thomas (2004) puts it: "Evaluation is not simply something done to the school context or reform intervention, but instead is something that is done for the school and its intervention efforts" (p. 6).

The next section present strategies for conducting culturally responsive evaluations as discussed in the 2002 and 2010 *National Science Foundation (NSF) User-Friendly Handbooks for Project Evaluation.*

The 2002 and 2010 User-Friendly Handbooks for Project Evaluation. The *NSF 2002 and 2010 User-Friendly Handbooks for Project Evaluation* were developed to provide program managers in NSF with practical guidebooks for the development and implementation of evaluations of educational programs offered by NSF. In 2010, the handbook was revised. Each guidebook provides information on various types of evaluations, methods, and strategies. In addition, one section in each handbook is devoted to culturally responsive evaluation. The sections were entitled in 2002 and 2010, respectively: a) strategies that address culturally responsive evaluation and, b) a guide to conducting culturally responsive evaluation.

Both sections provide information relevant to strategies, tools, and techniques that can be used in the development and implementation of culturally responsive evaluation. Written by three prominent scholars in the field the purpose of the guide is to familiarize evaluators with strategies relevant to culturally responsive evaluation. Further, the 2010 revised handbook provides additional strategies for conducting culturally responsive evaluation, and in addition provides illustrative examples from the field.

The sections highlight strategies that can be used in culturally responsive evaluation and are discussed for each phase of the evaluation. The diagram below illustrates the culturally responsive evaluation framework (Hopson, 2010). The purpose of the framework is to incorporate the cultural context of the program at every phase of the evaluation. The guide is divided into the nine stages of the evaluation process illustrated in the diagram below.

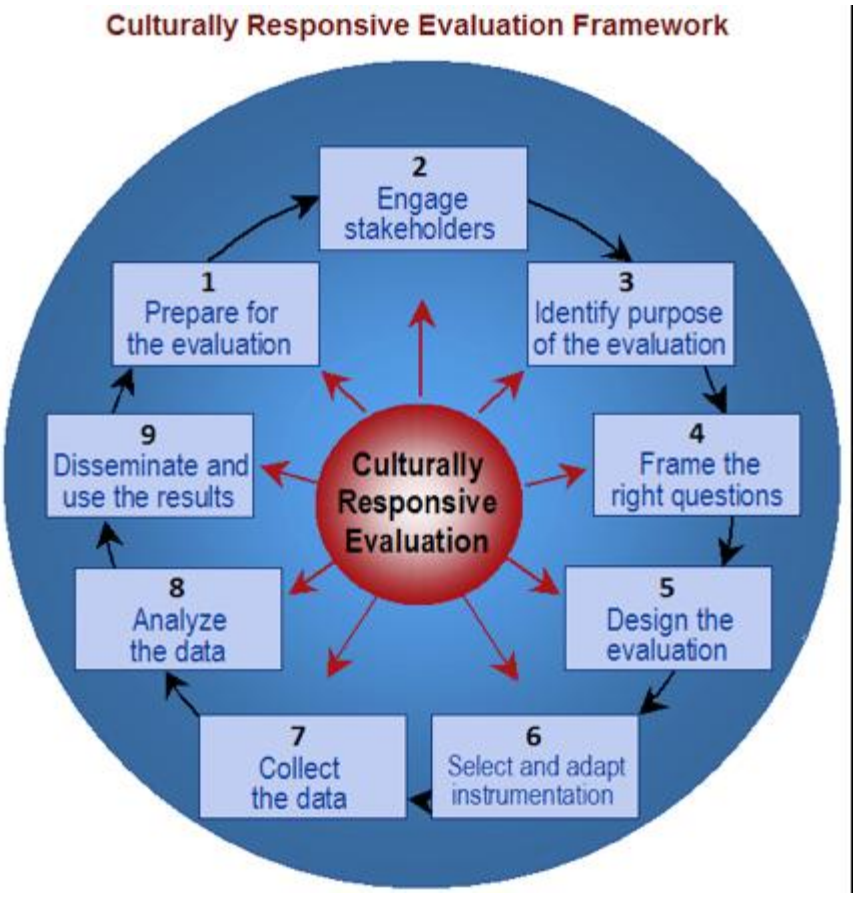


Figure 1. Culturally responsive evaluation framework.

More specifically, the section highlights key strategies for designing and implementing culturally responsive evaluation for each evaluation phase.

Preparing for the evaluation. Prior to the evaluation, a culturally responsive evaluator takes into account the cultural context of the program: "At the outset, the sociocultural context in which the programs or projects are based in must be taken into account" (Frierson, et al., 2002, p. 65). Evaluators can employ a multitude of strategies to become familiar with the program prior to the evaluation. For example, evaluators can conduct a historical analysis of the program to gain information relevant to its history and development. Evaluators can also spend a significant amount of time within the context of the program to obtain contextual information relevant to the

program's characteristics. Finally, evaluators can administer interviews with program staff, administers, and students to get first-hand information on the program.

A critical component in the preparation of a culturally responsive evaluation is the development of a multi-ethnic evaluation team. A multi-ethnic evaluation team increases the "chances of really hearing the voices of underrepresented students" (Stevens, 2000, as cited in Frierson, et al., 2002, p. 65). Multi-ethnic evaluators are also more likely to have a lived experience similar to the populations under study. As such, a shared lived experience increases the likelihood that an evaluator truly understands the culture of the participants and the cultural context of the program. In cases where multi-ethnic evaluators cannot be selected for participation in the evaluation team, evaluators should be carefully selected. Individuals for consideration should include those "who understand or who at least are clearly committed to being responsive to the culture context in which the project is based" (p. 65).

Engaging stakeholders. Culturally responsive evaluators take into account the "voices" of all the populations the program serves. Culturally responsive evaluation carefully identifies group representatives "of the populations the project serves, assuring that individuals from all sectors have chance for input" (Frierson, et al., 2002, p. 65). While there is a special emphasis on selecting stakeholders with less power (i.e. parents and students), culturally responsive evaluation targets all groups impacted in one way or another by the program (i.e. clients, program staff, administrators, community members, and schools). Stakeholder input in culturally responsive evaluation is important because stakeholders can provide "sound advice from the beginning (framing questions) to the end (disseminating the evaluation results)" of the evaluation (p. 65). Stakeholder input is therefore considered at every phase of the evaluation process. Evaluation stakeholder participation is particularly important in the interpretation and

dissemination of the evaluation findings. Furthermore, stakeholder participation in the evaluation ultimately leads to greater accuracy of the evaluation results.

Identifying the purpose(s) and intent of the evaluation. Evaluation is performed for various purposes: a) to improve a program's performance (formative) or b) to determine the extent to which a program has achieved its intended goals (summative). In both formative and summative evaluation, it is critically important that evaluators be explicit about the evaluation's purpose. Additionally, culturally responsive evaluators have extra responsibilities in both formative and summative evaluation. In formative evaluations, culturally responsive evaluators describe and explain program activities through a cultural lens. This approach requires that the evaluators are both sensitive to and respectful of the program environment. Additionally, in summative evaluation, "careful documentation of the implementation of program activities is critical to making sense" of the evaluation results (Frierson, et al., 2002, p. 66).

Framing the right questions. Framing the evaluation questions is an important task in any evaluation. The evaluation questions should be developed before the evaluation begins and in consultation with key evaluation stakeholders: "For an evaluation to be culturally responsive, it is critical that the questions of significant stakeholders have been heard and, where appropriate, addressed" (p. 67). Additionally, the evaluation questions should be appropriate for the program context, congruent with the program's cultural context: "Thus, an important key to successful evaluation is to ensure that the proper and appropriate evaluation questions have been framed" (p. 67), in the case of culturally responsive evaluation relevant to the program context.

After evaluation questions have been developed, the next step is to determine what type of data is necessary for answering the evaluation questions. In this phase of the evaluation, evaluators are confronted with the following question: "What will we accept as evidence when

we seek answers to our evaluative questions?" (p. 67). In culturally responsive evaluation, discussions about what is creditable evidence are facilitated democratically. As such, key evaluation groups are asked to come to the table and share their thoughts about what will count as credible evidence in a particular evaluation. Although discussions about acceptable evidence are not always easy to have, they are nonetheless "necessary" and should occur prior to the evaluation (p. 67).

Designing the evaluation. There are various types of evaluation designs (i.e. experimental, quasi-experimental, mixed methods, case study). Evaluation designs are selected based on how well they can address the evaluation questions. Thus, the evaluation design should be appropriate for the evaluation context, as well as aligned with the needs of the program. In culturally responsive evaluation, "comprehensive" evaluation designs are preferred. Comprehensive evaluation designs include both qualitative and quantitative methods. Qualitative and quantitative methods provide "data in a format that is different from the other, but that can also be complementary to other" (p. 68). Additionally, culturally responsive evaluations administer the methods of data collection, surveys, interviews, and observations at multiple times over the course of the evaluation (pre- and post evaluation). Furthermore, control or comparison groups can also be used when and if appropriate. This is not always possible in culturally responsive evaluation due to the complex nature of the programs under study; however, they should be incorporated into the evaluation if appropriate to the context of the evaluation to increase the rigor of the overall evaluation.

Selecting and adapting instrumentation. Evaluation instruments should be identified based on how appropriate they are for the particular evaluation context. The evaluation instruments should garner necessary information for answering the evaluation questions.

Furthermore, the evaluation instruments should be both reliable and valid. Thus, instruments that have "history" should be used, over instruments that have not been tested for either reliability or validity (Frierson, et al., 2010, p. 86). Existent instruments should be used with caution; in some cases will need to be normed for the target population. In addition, when working with populations in which English is not their first language, instruments should be translated. There are various ways to translate instruments; they include a) forward/back translation (FBT) and b) translation by committee (FBC) (Frierson, et al, 2010, p. 86)

Collecting data. Culturally responsive evaluation makes use of both qualitative and quantitative data. However, qualitative data is more common in culturally responsive evaluation. When collecting data, the evaluator should be attuned to the cultural context of the program. Who collects the data and how the data is collected impacts the validity and reliability of the data that is collected. Thus, the data should collected by someone who is culturally attuned to the program's context. Moreover, the data should be collected in culturally responsive ways. For example, the evaluation instruments should be culturally sensitive to the population under study in the evaluation (i.e. instruments, for example, surveys, should be written in Spanish if the evaluator is working with only a Spanish speaking population). When collecting data in culturally responsive evaluation, an evaluator who possesses a shared lived experience is more likely to obtain valid data. When an evaluator lacks a shared lived experience, program staff are encouraged to hire extra staff or teams that reflect the racial/ethnic composition of the program participants. It is not a *requirement* that the evaluator share the same racial/ethnic background; however, it is strongly encouraged and should be pursued when possible.

Analyzing the data. Data must be analyzed from a culturally responsive lens. For example, qualitative data should be analyzed with "considerable sensitivity to, and understanding

of, the cultural context in which the data was gathered" (Frierson, et. al, 2002, p. 70). In addition, how the data is interpreted, and what meanings are derived from it, is an important component of culturally responsive evaluation. Data analysis and interpretation often requires an evaluator to have a critical eye. It may also require the evaluator to conduct a critical examination of the cultural and contextual landscape of the program context. Additionally, a shared lived experience is also beneficial when analyzing and interpreting the evaluation results. Cultural nuances are more likely to be identified and interpreted accurately in the evaluator shares the same racial/ethnic background of the participants.

In addition, quantitative data should be analyzed in various ways. "Disaggregation of collected data is a procedure that has gained increased attention in education and the social sciences in general" (Frierson et al., 2010, p. 90). Disaggregation of data can be used, and also cross-tabulations to analyze the data in various ways. In addition, data can be disaggregated by race, ethnicity, and socioeconomic status.

Disseminating and utilizing the results. Evaluation results should be disseminated broadly and to multiple audiences. The evaluation should also be disseminated effectively. That is, the evaluation results should be disseminated "in manner that increases the likelihood that the results will be perceived as useful, and indeed, used" (p. 71).

Evaluation results should also be presented in an understandable manner and in ways from which multiple audiences can benefit (i.e. underrepresented populations). Evaluators can present the evaluation results in more than one way. Evaluators can report the results in formal presentations and reports, as well as in alternative representations such in the form of poems, skits, and dance.

Ethical considerations. In culturally responsive evaluation, evaluators are ethical and are bound to two types of ethics when conducting evaluations. That is, procedural ethnics which include those required by Institutional Review Boards, and relational ethnics which include "mutual respect, dignity, and the connectedness between the researcher and the researched and between the researchers and the communities in which they live and work" (Frierson et al., 2010, p. 92). Of the two types of ethics listed above, culturally responsive evaluation is more concerned with the relational ethics because culturally responsive evaluations typical occur in program contexts in which there are underserved and marginalized communities. In evaluation contexts where there are clear power imbalances, evaluators respect and honor the culture of the communities and are "mindful on building relationships of trust and mutual respect" (p. 93).

The *2002 and 2010 User-Friendly Handbooks for Project Evaluation* provide evaluators with a host of strategies for conducting culturally responsive evaluation. To date, these are the two most comprehensive guides for culturally responsive evaluation. In the future, it is likely that the field will provide more practical guidebooks for conducting culturally responsive evaluations.

Examples of Culturally Responsive Evaluation

The present section examines two case study examples of culturally responsive evaluation. The present evaluations address the benefits and challenges of designing and implementing culturally responsive evaluation. In addition, they address how culturally responsive evaluation can be applied in practice. The following case studies of culturally responsive evaluations were conducted in two program contexts: a) the Family, School, and Community Partnership Program (FSCPP) and; b) the Ninth Grade Career Breakfast Club Program.

LaPoint & Jackson (2004) evaluated the co-construction, a component of the TD evaluation framework of an academic support program for secondary schools and students. The goal of (FSCPP) was to improve the academic achievement and social competence of African American high school students in low-income urban schools. To examine the co-construction of FSCPP, the major themes of the following evaluation approaches/models were applied to the development, implementation, and evaluation of CRESPAR. The four models included: CRESPAR's Talent Development School Reform Model, the Talent Development (TD) evaluation framework, the National Network of Partnership Schools (NNPS), and participatory action research (PAR). In collaboration with school stakeholders (students, teachers, staff, family, and community members), the five key components of the FSCPP program were developed: a) the Family Resource Center (FRC), b) the High School Action Team (HSAT), the Talent Development Attendance Program (TDAP), the Talent Development Student Team (TDST), and the Newsletter Communication Network (NCN).

The FSCPP evaluation demonstrates how meaningful engagement was a key component of the TD evaluation. TD evaluators engaged school stakeholders on multiple levels of program development, implementation, and evaluation. Evaluators utilized four strategies to involve key stakeholders groups in the evaluation process. First, program staff engaged stakeholders early in the evaluation by hosting a meeting with the Local School Restructuring Team (LSRT). At the meeting, members of the community were provided with information on the development, implementation, and evaluation of FSCPP. In addition, "TDSSP staff presented information on co-constructing an evidence-based program" (LaPoint & Jackson, 2004, p. 29). Second, an action team composed of school stakeholders was set to guide the multiple initiatives of FSCPP. The High School Action team members "planned, guided, and monitored the implementation of

FSCPP interventions and facilitated stakeholder involvement in educational programming and school wide committee" (p. 28). Third, a series of meetings were held with individuals and groups to collect data about school climate, organization, and culture. Finally, school members were provided the opportunity to serve as assistant evaluators, which led to greater stakeholder involvement in the evaluation.

Additionally, Manswell-Butty, et al., (2004) examined how a culturally responsive approach was used by TD evaluators to study a CRESPAR program. The Ninth Grade Career Breakfast Club Program seeks to assist high school students with career exploration and planning by using the Holland Self-Directed Search Self-Assessment. In this study, Manswell-Butty, et al., (2004) focused specifically on the extent to which the evaluation of the Ninth Grade Career Breakfast Club met the culturally responsive evaluation requirements outlined by Frierson, et al., (2002): (a) preparing for the evaluation, (b) engaging stakeholders, (c) identifying the purpose(s) and intent of the evaluation, (d) framing the right questions, (e) designing the evaluation, (f) selecting and adapting instrumentation, (g) collecting data, (h) analyzing data, and (i) disseminating and utilizing the results.

To evaluate the Ninth Grade Career Breakfast Club Program, TD staff used a culturally responsive evaluation approach to meaningfully engage stakeholders in the evaluation process. Attempts to engage stakeholders in all phases of the evaluation as outlined by Frierson, Hood, and Hughes included: a series of meetings with stakeholders to obtain information about the program, the students, and the environment of the program, the "identification" of key stakeholders before the evaluation to serve as liaisons between the TD evaluators and program staff, and the implementation of both a summative and formative evaluation to provide program staff with information about student progress at various stages of the evaluation process. In

relation to the design and implementation of the evaluation, TD evaluators also used culturally responsive strategies to gain the support of school administrators. These strategies included: (1) selecting instruments that were culturally appropriate; (2) using multiple data collection instruments (qualitative and quantitative) to answer the evaluation questions; (3) collecting data from a variety of sources to obtain data reliability (pre-test, post-test, and a comparison group); and (4) analyzing data in multiples ways (by race, gender, and age) (Manswell-Butty et. al, 2004). After reviewing the evaluation, the authors demonstrated the benefits of applying a culturally responsive evaluation approach to the Ninth Grade Career Breakfast Program. Moreover, the evaluation met the requirements as outlined by Frierson, et al., in the *2002 User-Friendly Manual for Project Evaluation.*

Challenges to Culturally Responsive Evaluation

There are a number of challenges to designing and implementing culturally responsive evaluation. The challenges of conducting culturally responsive evaluation include but are not limited to: (a) a lack of representation among evaluation stakeholder groups (Greene, 2000; Thomas 2004; Ryan, et al., 2007); (b) a shortage of staff to fulfill all evaluator roles (Manswell-Butty, et al., 2004); (c) competing agendas between schools, districts, and communities (Thomas, 2004); (d) the complex and dynamic nature of urban school settings (LaPoint & Jackson, 2004); (e) limited time and/or staff to complete all tasks necessary in culturally responsive evaluation; and (f) limited money or resources. The two most common challenges to implementing culturally responsive evaluation are (1) labor and time intensive, (2) extensive costs (Hood, 2011, personal conversation).

The next section discusses some of the challenges related to the practice of culturally responsive evaluation. Two case study examples are provided to examine the types of challenges

evaluators experience when conducting culturally responsive evaluation, as well as the ways in which these challenges limit an evaluator's ability to plan, design, or carry out a culturally responsive evaluation. The first study is an evaluation of the Family, School, and Community Partnership Program (FSCPP) (LaPoint & Jackson, 2004). The second study is an evaluation of Ninth-Grade Career Breakfast Club Program (Manswell-Butty, et al., 2004). These studies were discussed in the section above regarding the benefits to culturally responsive evaluation.

The FSCPP evaluators reported a variety of challenges to implementing culturally responsive evaluation. The challenges were present at various levels: program development, implementation, and evaluation. First, school and district agendas did not support the goals of the evaluation. "Differing, competing, and ever-changing school-and district-level agendas during and after co-constructed partnerships agreements" had a negative impact on the evaluation (LaPoint & Jackson, 2004, p. 33). Different political agendas held by school personnel limited the capacity of the evaluation. Second, high turnover rates and new program personnel created instability for district level administrations. High turnover is unfortunately a prominent challenge urban schools face. Third, participants had negative experiences with prior school programs, which led them to not want to participate in the program and evaluation (LaPoint & Jackson, 2004).

The Ninth-Grade Career Breakfast Club Program presented a variety of challenges to the implementation of a culturally responsive evaluation. The first challenge was the intensity of the labor needed to develop and implement a large-scale evaluation. The Ninth-Grade Career Breakfast Club Program was composed of several different components, which made the work of the evaluators' very labor intensive. Also, there was very limited staff, so evaluators had to play

multiple roles, from “developer to evaluator to curriculum specialist” (Manswell-Butty, et.al., 2004, p. 45).

Benefits to Culturally Responsive Evaluation

While there are challenges, there are also a number of benefits to culturally responsive evaluation. To highlight some of the benefits two examples of culturally responsive evaluation will be presented. Overall, culturally responsive evaluation and other similar evaluation approaches/paradigms that endeavor to be culturally responsive have “helped in documenting program processes, effectiveness, and outcomes, as well as providing specific benefits to projects and their stakeholders” (Maswell-Butty, et al., 2004, p. 37). The first study deals with an evaluation of an Upward Bound program (Zulli & Frierson, 2004), and the second study focuses on evaluations of STEM programs for underrepresented minority students (Mertens & Hopson, 2006).

An evaluation conducted by Zulli & Frierson (2004) demonstrated the benefits of implementing a culturally responsive evaluation approach to an evaluation of the Upward Bound Program. Upward Bound prepares underrepresented high school students for college through intensive coursework. The present evaluation documented the experiences of the program developers in their attempt to apply a culturally responsive approach to develop and implement an evaluation of the Upward Bound Program. In this study, culturally responsive evaluation was used to examine cultural factors within the Upward Bound program. Zulli & Frierson (2004) noted: “The example set by this Upward Bound program’s developers was not merely taking culture into account but going so far as to proactively build in what could be viewed as cultural responsiveness in program design” (p. 83).The results from the evaluation demonstrated the benefits of addressing cultural factors within the Upward Bound program (Zulli & Frierson,

2004). These benefits occurred primarily in the staffing of the program that had a similar, if not the same, racial and ethnic background as the participants. The benefits of having a staff who shared similar experiences to the participants included students having an increased sense of belonging and respect toward the other program participants and staff. In addition, by ensuring that program staff members were culturally competent on the diversity of issues related to this population, the program developers were able to relate to the students on multiple levels. As a result, many of the students expressed that the program staff had in some ways served as role models for them after the program ended. Moreover, some of the benefits associated with the program also were a result of program staff being committed to the goals of the program and having a sincere interest in helping the participants of the program.

Another example of a study that applied a cultural responsive evaluation approach is a study on STEM programs (Mertens & Hopson, 2006). This evaluation focused specifically on implementing various components of cultural competency and transformative evaluation to examine issues related to STEM programming for minority students. Through these lenses, the researchers were able to examine the role that both diversity and culture played in program efforts targeted at increasing the representation of ethnic minorities in STEM fields. The benefits of employing transformative and cultural competency evaluation included: (1) allowing the researchers to look at the complexity of issues related to minority student experiences within STEM programs; (2) providing a basis for an examination these experiences from an array of perspectives, similar to that of a prism, that reflects ever changing patterns; and (3) allowing the researchers to take into account multiple dimensions of culture/context with relation to issues of race, oppression, and discrimination (Mertens & Hopson, 2006).

Future considerations for culturally responsive evaluation

The present section highlights themes within recent literature relevant to culturally responsive evaluation. The issues discussed here are presented because their importance within the field presently, but also because they are somewhat recent developments. Future research on culturally responsive evaluation should take into account theories of language, race, and gender.

Language: Research demonstrates the importance of the role of language within the field of evaluation (Hopson, 2000, Patton, 2000; Madison, 2000; Hopson, Lucas, & Peterson, 2000). Evaluators who employ culturally responsive evaluation should take into account the many ways in which language matters, and how it is related to concepts of culture and context. Patton (2000) states, "The evaluation language we choose and use, consciously or unconsciously, necessarily and inherently shapes perceptions, defines, 'reality,' and affects mutual understanding" (p. 5). As such, how concepts are defined and used, such as culture, race, and power, can have adverse effects on the ways in which programs are represented and participants are classified. As Madison (2000) suggested, "labeling is a sociopolitical process used by those in power to establish themselves as the norm" (p. 19). This situation is especially the case for programs that serve a high population of ethnic minorities and disadvantaged populations. Furthermore, Madison (2000) examined the ways in which the term "at-risk" was used in an intervention program aimed at providing services to youth considered to be at "at-risk." Madison (2000) reported the following finding from this study: "In this context, at-risk youth was used to describe a particular kind of youth and to sort youth by ethnic and social class group identify. Thus, the term at-risk as used in this context places emphasis on the youth as a category, rather than the risk factors" (p. 10).

Race. How issues of race are considered within educational evaluation is an important issue which should be addressed in the context of culturally responsive evaluation. Critical race

theory (CRT) examines the ways in which race and racism are embedded within structures of society. CRT places race at the forefront of the analysis to examine the ways in which racism functions in everyday society, especially in theories of color-blindness (Parker, 2004). As an analytic tool, culturally responsive evaluation can benefit from a theory that takes into account the role of race, specifically how racism is embedded within educational institutions. A CRT framework provides evaluators with the ability to examine issues of race at a deeper level. Parker (2004) states, "It is here that CRT could be potentially useful as an overall framework for future evaluation studies of this type, given how it has evolved as a methodology perspective on race, law, and educational policy" (p. 89). Overall, culturally responsive evaluations would benefit by incorporating tenants of CRT.

Gender. Feminist theory has various implications for culturally responsive evaluation theory (Hood & Cassaro, 2002; Pillow; 2002, Mertens, 1999). First, feminist theory fits under the genre of critical theory. It seeks to "challenge assumptions of universal concepts and acknowledges that 'reality' is socially constructed" (Hood & Cassaro, 2002, p. 28). To do this, feminist theory takes into account multiple levels of oppression as influenced by differences of sex/gender, race, social class, sexual orientation/preference, and reprise (Hood & Cassaro, 2002). As a methodological approach, feminist theory engages the issue of "difference." Therefore, feminist theory can be applied to culturally responsive evaluation in order to gain a deeper understanding of some of the issues discussed within feminist theory.

Moreover, feminist theory argues that the race, gender, class, and sexual orientation of the evaluator matters (Hood & Cassaro, 2002). This approach supports having a shared lived experience in culturally responsive evaluation. It is believed that the identity of the evaluator impacts the processes of data collection, analysis, and interpretation. Evaluators who have a

shared background with participants are more likely to meet the needs of the population served by the evaluation. When there is a lack of congruency between the evaluators and the participant's background, an evaluation may be "imbued with racist, heterosexist, or classist attitudes" (p. 33) and will yield "inaccurate results" (p. 33). This approach demonstrates how feminist theory shares some of the same principles of culturally responsive evaluation. As such, these similarities should be considered in future studies examining culturally responsive evaluation.

The need for more culturally responsive evaluation

The importance of culture and context cannot be ignored in evaluations that seek to meet the needs of marginalized groups. Significantly, as the number of ethnic minorities increases, it becomes paramount that evaluators adopt evaluation strategies and techniques that are culturally responsive to various minority populations in the US (i.e., Latinas/os, African Americans, Asian Americans, and Native Americans). It is within this context that culturally responsive evaluation provides a template in which evaluators can better understand the ways in which culture and context influence program design, implementation, and evaluation. Moreover, to ensure that evaluators make accurate judgments about program quality, it is important that evaluation is attuned to the cultural context of the program and the participants.

In the 21st century, the presence of culturally responsive evaluation is not only timely but also necessary. In light of the recent shifts in the nation's racial, ethnic, and social demographics, evaluators should employ evaluation approaches that are culturally responsive to the context of the evaluation. Evaluators who support evaluation approaches that are culturally responsive argue that evaluation is not only important, but also necessary for conducting good evaluations. Hood, Hopson, and Frierson (2005) purport: "Without the nuanced consideration of cultural

context in evaluations conducted within communities of color and/or poverty there can be no good evaluation" (p. 1). Therefore, for an evaluation to be a good evaluation, the cultural context of the program should be a key component of the evaluation.

Although culturally responsive evaluation should seem like a viable evaluation strategy for the evaluation community at large, unfortunately it is not. Culturally responsive evaluation has been challenged on many fronts. Primarily, evaluators have questioned the validity of the evaluation approach. Some argue that evaluation should be as objective as possible; therefore it should not include the cultural context in the evaluation because it makes it subjective. Moreover, others have argued that it is not the evaluators' duty to respond to the culture of the program (Frierson, et al., 2002; Manswell-Butty, et al., 2004)

There are many benefits of culturally responsive evaluation. Primarily, an increased understanding of the factors that affect a program's design, implementation, and evaluation. The benefits of culturally responsive evaluation certainly outweigh the challenges. There are, however, many challenges to conducting culturally responsive evaluation. These challenges primarily include 1) lack of resources and 2) lack of time. However, in situations in which culturally responsive evaluation is an appropriate evaluation approach for the context, evaluators should explore culturally responsive tools, strategies, and techniques for implementing a culturally responsive evaluation. If evaluators continue to ignore culture as a significant component of an evaluation, then the evaluation field and evaluators alike will continue to do a disservice to programs and their constituents. As a means to increase the field's knowledge on the practice of culturally responsive evaluation, this dissertation examines how a team of evaluators planned, designed, and implemented a culturally responsive evaluation.

CHAPTER 3

Methodology

This chapter outlines the methodology for this study, which consists of the research design, methods, data sources, instruments, and data analysis. The chapter begins with a discussion on the researcher's role and reviews the research objectives.

Researcher's Role

This study was conducted retrospectively, three years after the last evaluation of the G.A.M.E.S. program was completed. I was a participant of the evaluation, in the capacity of evaluator, for two of the three years the evaluation took place. As a member of the evaluation team, my responsibilities were to: (a) update the evaluation instruments; (b) collect and analyze the data; (c) write the evaluation reports; and (d) present the results to program stakeholders (i.e., program staff and administrators) orally and in written reports. My participation in the evaluation was limited to the second and third years. Therefore, I did not participate the first year the evaluation took place. Thus, I did not play a role in the preparatory phases of the evaluation. Further, I was not involved in the selection of the evaluation questions, purpose, design, methods, and instruments. My participation in years two and three of the evaluation consisted primarily of data collection, analysis, and reporting.

Because of my role of evaluator in the last two years of the evaluation, I recognize that there are certain limitations which have emerged from the study due to my insider researcher status. In the last section of this chapter, I discuss the limitations I encountered and how I countered them through the use of various techniques to reduce researcher bias. In addition to limitations, there are also advantages to being an insider researcher. For example, because I was

involved in the evaluation, I had first-hand knowledge of the program and participants. In addition, I had greater ability to make accurate judgments about the value, merit, and worth of the program because I was directly involved in the evaluation. The benefits of being an insider researcher are also discussed in this section.

Research Objectives

The primary objective of this study is to investigate the practice of culturally responsive evaluation. For example, what are the key components of culturally responsive evaluation and what does culturally responsive evaluation look like in practice? A secondary objective of this study is to examine how culturally responsive evaluations are conducted. For example, how are culturally responsive evaluations designed and/or implemented, and what types of evaluation strategies are used to develop a culturally responsive evaluation. Lastly, the purpose of this study is to explore the benefits and challenges of utilizing culturally responsive evaluation. For example, what obstacles do evaluators face when employing culturally responsive evaluation and/or what benefits do programs and participants experience from the use of culturally responsive evaluation? To achieve the research goals set out for this study, the primary research question in the case study is: *How do evaluators plan, design, and implement culturally responsive evaluations?*

Next, the study also addresses subsidiary questions, whose answers helped to inform the main research question:

1. What was the rationale for using a culturally responsive evaluation approach?
2. What strategies were used for incorporating the cultural context of the program in the evaluation?

3. What strategies were used for incorporating the culture of the participants in the evaluation?
4. What strengths and challenges did the evaluators face in conducting a culturally responsive evaluation?

The next section provides the context for the G.A.M.E.S evaluation by providing a description of the G.A.M.E.S. program and program participants. In Chapter 4, a more detailed description of the program context is provided.

G.A.M.E.S. Summer Camp

Program description. G.A.M.E.S. is an annual, week-long camp designed to give academically talented middle-school-aged girls an opportunity to explore engineering and scientific fields through demonstrations, classroom presentations, hands-on activities, and contact with women in these technical fields. The G.A.M.E.S. summer program was sponsored by the following organizations: Abbott Laboratories; Caterpillar; John Deere; Motorola; Shell Oil Company; ExxonMobil; the Center for Nanoscale Chemical-Electrical-Mechanical Manufacturing System; the Center for Excellence in Airport Technology; the Center for Preparedness, Response, and Recovery; along with several STEM departments on the campus of the University of Illinois at Urbana-Champaign (UIUC) (www.illinois.edu). In addition, G.A.M.E.S. was organized and sponsored by the Women in Engineering (WIE) Program at the College of Engineering at UIUC.

G.A.M.E.S. provided program participants with a variety of activities to learn and engage with various STEM fields. The primary goal of the activities was to expose and introduce program participants to careers in engineering, science and math. The activities included but were not limited to: (a) working on a team project; (b) learning about various engineering, math,

and science disciplines; (c) participating in hands-on demonstrations; and (d) meeting other young women interested in math and science (i.e., undergraduate and graduate students).

G.A.M.E.S. took place on the UIUC campus, and program participants lived with camp counselors (i.e., female undergraduate students majoring in STEM) in campus dorms. Four programs were available to the participants: 1) Structures, 2) Computer Science, 3) Bioimaging, and 4) Bioengineering/Chemical Engineering. Each program offered various activities relevant to their respective STEM fields. The participants were assigned to a program based on their interest and age.

Program participants. Program participants included girls entering grades 6 through 9. Program participants were recruited through several means. Brochures were emailed/mailed to girls who had attended G.A.M.E.S. in the past and to those who had participated in academic enrichment programs around the state. WIE personnel also recruited at schools in the surrounding areas (i.e., Champaign and Urbana), as well as through social networking, contacts with faculty and staff at the university, and via the internet (i.e., program and college website, wiki pages, and Facebook). In addition, special attention was placed by WIE personnel to recruit students from ethnically/racially and economically diverse backgrounds. To ensure the participation among these underrepresented students, WIE offered scholarships to students who were unable to pay the program fee.

The next section provides a description of the evaluation of the G.A.M.E.S. summer camp. The section addresses the evaluation's background, purpose, approach, design and methods. More detailed descriptions of the evaluation are provided in Chapter 4.

G.A.M.E.S. Evaluation

Background. The evaluation was not required by the College of Engineering, UIUC, or any outside agency. However, Women in Engineering (WIE) staff members expressed an interest in formally evaluating the program for the purposes of learning about the program, improving the program's design and implementation, and disseminating information about the program to others in the engineering education community.

Purpose. The primary purpose of the evaluation was to provide evaluative feedback to program administrators and staff regarding (a) the quality and appropriateness of program content; (b) the quality, effectiveness, and appropriateness of program pedagogy; and (c) the quality and meaningfulness of the program experience for its participants. Further, the purpose of the evaluation was to enhance educators' understanding of the program's characteristics and features as designed and implemented and to provide information useful for program improvement.

A secondary purpose of the evaluation was to address issues related to the reach, meaningfulness, and benefits of the program to diverse kinds of children, again for the purposes of enhancing program understanding and improvement. Furthermore, the evaluation was comprehensive in scope — addressing issues of project design, context, implementation, and outcomes, as well as surfacing values claims for reflection and critique.

Approach. An Educative Values-Engaged (EVEN) evaluation approach was employed in the evaluation (See Appendix A for a brief overview of the EVEN approach). The EVEN approach defines STEM program quality at the intersection of content, pedagogy, and equity and is inclusive of multiple perspectives and interests in STEM education (Greene, et al., 2006). Further, the EVEN approach addresses program quality from four vantage points: (1) the quality of the project content and instructional design for learners in the contexts served; (2) the

contextual relevance and power of the project design; (3) the advancement of the interests of under-represented and under-served groups; and (4) the quality of project implementation and outcomes.

Design and methods. The evaluation used a mixed-methods design in order to generate a comprehensive, in-depth understanding of the program of interest. Methods used included surveys, interviews, observations, and document analysis. First, all programs completed brief-pre- and post-surveys before and after the camp. The questionnaires focused on participants' experiences with G.A.M.E.S. summer camp. Second, a sample of camp participants and parents were interviewed to gather information on the perceptions of the quality and benefits of the G.A.M.E.S. summer camp. Third, observations were conducted to capture a descriptive record of the structure, content, and activities of the program and the characteristics of the students' participation in them. Finally, lesson plans and project manuals for the campers were collected and analyzed for descriptive information about program design, recruitment, participant selection, and key characteristics of program participants.

The next section describes the methodology used in this study. Individual descriptions of the research design, data collection methods and phases, data analysis, and limitations of the study are provided.

Methodology

Research design. This study is an *instrumental case study* (Stake, 1995) of the practice of culturally responsive evaluation. In an instrumental case study "a particular case is examined to provide insight into an issue or refinement of a theory" (p. 237). By employing the instrumental case study design, I used one particular case to examine the larger issue of the practice of culturally responsive evaluation. More specifically, the case played a supportive role

in the facilitation of knowledge on how evaluators plan, design, and implement culturally responsive evaluations. Therefore, in an instrumental case study, "The case is of secondary interest, it plays a supportive role, and it facilitates our understanding of something else" (p. 237). In this study, the G.A.M.E.S. evaluation was of secondary interest to the larger issue of culturally responsive evaluation and was studied in-depth in order to gain knowledge on culturally responsive evaluation.

The instrumental case study design was selected for this study because it is an appropriate research approach for this study. First, the goals of this research project are aligned with the goals of instrumental case study. For example, the present research study seeks to address research questions which begin with why and how – and in case study research, why – and how questions are appropriately addressed in case studies (Yin, 1994). Further, the purpose of the present research study is to describe, explain, and understand particular phenomena, and in case study research, the purpose is to examine and describe in-depth issues or phenomena. Further, the instrumental case study design was selected because it helped me to answer my research questions. First, the use of instrumental case study design allowed me to study one particular case to gain insight into a larger issue of interest: the practice of culturally responsive evaluation. Further, the present study aimed to highlight the particularity and complexity of a single case study to develop understanding of the issue of interest (Stake, 1995).

Case selection. The evaluation of the G.A.M.E.S. summer camp was selected to study culturally responsive evaluation for a number of reasons. First, the selection of the case was practical. Due to my own previous experience with the evaluation in the capacity of evaluator, I had extensive knowledge of the program and program participants. In addition, I was familiar with the context in which the evaluation took place. Therefore, my previous experience in the

evaluation gave me insight into the evaluation, and also granted me access to conduct this study. For example, both the client and evaluators expressed interest in the study. In addition, my interest in culturally responsive evaluation and STEM evaluation also led me to conduct this study.

The evaluation was also selected because it was appropriate. First, the purpose of the evaluation was to attend to issues of culture and context. Additionally, the evaluation included dimensions of culturally responsive evaluation which were apparent in both the design and implementation of the evaluation. Further, instances of cultural responsive evaluation were present in the evaluation documents. Additionally, the evaluation was designed to achieve cultural responsiveness. For example, the evaluation was designed and implemented though the use of a variety of culturally responsive evaluation strategies. The evaluation strategies were culturally responsive; aimed to take into account the cultural context of the program.

Further, the evaluation was selected because of its nature. The evaluation was large-scale and longitudinal. First, it was large-scale because a significant amount of information was collected on the participants, and others involved in the evaluation via multiple data collection methods. In addition, a large number of individuals participated in the evaluation in the role of student, faculty, staff, administrator, undergraduate and graduate student (n=250). Further, the evaluation was longitudinal, taking place over a three year period. Three separate evaluations were conducted in the summers of 2007, 2008, and 2009. Data collection took place over a two-week span and evaluation activities spanned into the fall semester (i.e. evaluation presentations).

Data Collection Methods

Multiple data collection methods were used to garner data relevant to the research purpose. The methods used in this study included: (a) observation review, (b) document review,

and (c) formal and informal in-depth interviews. The use of qualitative data collection methods help to garner rich in-depth data to answer the main research question in this study: *How do evaluators plan, design, and implement culturally responsive evaluations?* The qualitative data garnered also helped to achieve the multi-fold purpose of the study, which was to describe, explain, and understand how evaluators prepare for, develop, and implement culturally responsive evaluations. A description of each data collection method is provided below.

Observation review: The first data collection method was observation review. The purpose of observation review was to gather information relevant to how the evaluation was carried out through the use of a culturally responsive evaluation approach. To do this, I reviewed two forms of observations: 1) participant fields notes and 2) observations of program activities. Participant field notes were notes taken by the evaluation team relevant to the evaluation process. In addition, I also reviewed minutes from evaluation team meetings and client evaluation meetings and electronic documents pertinent to the evaluation, including emails and files related to IRB procedures. Lastly, notes and presentation slides from evaluation presentations given to the client and program staff were also reviewed (See Appendix B for evaluation presentation slides).

In addition to participant field notes, I also reviewed observation notes taken by the evaluation team of program activities. These observations were taken via an observation protocol developed by the evaluation team to record and document participants' engagement in the activities (See Appendix C for observation protocol). Observation notes were taken on the following categories:

1. Description of context: a description of the context/setting, including its physical features, how teacher and student are arranged in the room and the general classroom atmosphere, using as much descriptive language as possible.
2. Descriptive account of program activities: a description of the key activities (e.g., presentations, group activities, video, lectures, hands-on, paper / pencil tasks, or outdoor activities), instruction (e.g., interactive, directive, amount of response time provided to learners, pace of instruction, use of learners' prior knowledge, questioning, recall, incentives, etc.), and use of materials (e.g. visuals, manipulatives)
3. Description account of learners' participation: a description of 4-5 randomly selected program participants during a small group or individual activity focused on (a) the character of the learners' attention, engagement, and enthusiasm, as well as frustration, boredom, or disinterest; (b) students' interactions with one another, the instructor, and the materials; and (c) any other event or response that relates to the students' experience.

Document review. The second data collection method was document review. First, a document review of evaluation materials was performed. Evaluation materials were reviewed to gather information related to the use of culturally responsive evaluation. Materials reviewed and analyzed included: a) evaluation plans, b) evaluation instruments, c) evaluation presentations, d) evaluation reports, and e) evaluation agendas. Of the evaluation materials listed, the majority of the document review efforts were focused on the evaluation reports. The evaluation reports included information relevant to the program, evaluation context, evaluation purpose, evaluation audience, key evaluation questions, criteria for judging program quality, evaluation design and methods, program participants, findings, and summary of findings. Therefore, the evaluation reports provided rich data on the evaluation.

Second, a document review of program materials was performed. Program materials were reviewed and analyzed to learn about the program, program context, and participants. Program materials reviewed included: a) participant manuals, b) staff manuals, c) promotional materials (i.e. website and brochures), and d) archival records (i.e. student demographic databases). First, participant manuals garnered dates related to the camp schedules, camp activities, final projects, program reading materials and activity worksheets. Second, staff manuals garnered data on the participants, activities, and program guidelines. Third, promotional materials garnered data on the camp, application procedures, activities, schedules, sponsors, staff opportunities for employment, and eligibility. Lastly, the archival records garnered information on the participants, including demographics.

Interviews. The last data collection method was interviews. The purpose of the interviews was to gather information relevant to the use of culturally responsive evaluation. First, informal interviews (n=25) were administered throughout the evaluation. Informal interviews were administered with evaluation team members, program administrators and staff. Second, formal in-depth interviews (n=3) were administered with the client of the evaluation and the other two evaluation members.

Interview protocols were developed for the interviews and tailored to the interviewee's role in the evaluation (i.e. evaluator or program staff). Both interview protocols included questions pertaining to culturally responsive evaluation and the use of a culturally responsive evaluation approach in the G.A.M.E.S. evaluation (see Appendices D and E for evaluator and administrator interview protocols, respectively). The first set of interview questions probed the interviewees' general understanding of the practice of culturally responsive evaluation. The second set of questions probed the interviewees' use of culturally responsive evaluation in the

G.A.M.E.S. evaluation. Both interview protocols included similar questions, and in some cases the same questions. Further, the purpose of including the same questions in both interview protocols was to increase the reliability of the study's inferences. Also, the inclusion of the same questions helped to garner different perspectives on the same topics covered in the interviews.

As mentioned above, multiple data collection methods were used to garner a comprehensive understanding of the ways in which the G.A.M.E.S. evaluation was culturally responsive. For example, the review of observations notes provided me with information relevant to the evaluation team's use of culturally responsive evaluation. Second, the document review provided me with examples and instances in which culturally responsive evaluation was employed throughout the evaluation. Lastly, the interviews provided me with first-hand information about how the evaluation was culturally responsive from the perspective of the evaluators and the evaluation client.

In addition to a comprehensive understanding, multiple data collection methods were also used to achieve data triangulation. Data triangulation (Denzin, 1978) examines the consistency of data across multiple methods. First, I compared and contrasted the data gathered from each method to identify similarities and differences in interviewees' responses and data provided via document review and analysis. Secondly, I reviewed the data to ensure that it was consistent across the three data collection methods. The use of data triangulation increased the validity of the results through cross-validation: comparing the data across methods.

Data Collection Phases

As depicted in Table 1 below, data collection in this study occurred in three phases (See Table 1). First, Phase 1 of data collection took place in the summers of 2008 and 2009. Data collection in Phase 1 included observations of the evaluation process and observations of

program activities. This was the phase in which I was an evaluator of the program. Second, Phase 2 of data collection took place in the summer of 2011. Data collection in Phase 2 included document review of program materials and evaluation documents. This phase included a collection of documents related to the evaluation and program. Data collection in Phase 3 took place in the fall of 2011. This phase included the administration of three formal, in-depth interviews. Interviews were done with the client of the evaluation and the other two evaluators. Therefore, data collection in this study occurred over a span of four years. First, as a participant in the evaluation and then as a researcher for this study.

Table 1

Data Collection Process

Phase	Date	Method
1	Summer 2008 & 2009	Observations
2	Summer 2011	Document Review
3	Fall 2011	Formal Interviews

Data Analysis

First, theme analysis was used to analyze qualitative data from all three methods. Data garnered in Phase 1 of the data collection process were reviewed. This included observation notes from the evaluation process and program activities. Participant field notes of the evaluation process were garnered from various evaluation materials (i.e. evaluation meeting minutes), and these were reviewed and analyzed to identify common themes or differences among the data. In addition, notes from observations of the program activities were reviewed and analyzed, and themes among the data were also identified.

For data garnered in Phase 2 of the data collection process, the qualitative data was analyzed thematically. All evaluation materials were reviewed and themes were identified. The evaluation reports were analyzed more systematically. Based on systematic review of the literature on culturally responsive evaluation, a standardized rubric was developed (See Appendix F for rubric). In this literature review, I identified the most frequently cited dimensions of culturally responsive evaluation, and I included these in the rubric. When reviewing the evaluation, the rubric guided my analysis of the evaluation reports, for example which dimensions of culturally responsive evaluation to look for. Further, the rubric required me to determine the extent to which the identified dimensions in the literature on culturally responsive evaluation were included or demonstrated in the evaluation reports. Using this rubric, I rated each dimension on a four-point criterion from absent (rating 0) to a great extent (rating 3).

1. Absent (rating 0): The evaluation did not address this dimension of culturally responsive evaluation;
2. To some extent (rating 1): The evaluation addressed to some extent this dimension of culturally responsive evaluation;
3. To a moderate extent (rating 2): The evaluation addressed to a moderate extent this dimension of culturally responsive evaluation; or
4. To a great extent (rating 3): The evaluation addressed to a great extent this dimension of culturally responsive evaluation.

Further, the rubric also required me to provide a written description of how each of these dimensions was addressed in the evaluation reports. Thus, this rubric produced data that was both quantitative (number ratings) and qualitative (written description) in nature.

Finally, for the qualitative data gathered from the three in-depth interviews in Phase 3 of the data collection phase, theme analysis was also utilized. For each interview, a summary transcription of the field notes and an audio transcription were produced, and themes within both the summary transcription and field notes were identified. In addition, the themes within the summary transcriptions and audio recording were compared. In some cases, similarities and differences across the respondents' answers emerged and these were explored in the data analysis phase. Further, the data garnered in the second phase of data collection was followed up on during the interviews. Thus, the interviews served to either confirm or to negate the data garnered in the previous data collection phases.

Data Quality

High quality data are essential for any research project. To ensure the collection of high quality data, rigorous criteria for judging data quality are established. Importantly, data in research studies should be both reliable and valid. The criteria for judging data quality in this study included Guba and Lincoln's (1985) criteria for judging qualitative research. The two criteria that were used to establish trustworthiness included: (a) credibility and (b) conformability.

Credibility. The results of a study are credible when the study measures what it set out to measure or test (Shenton, 2004). To meet the credibility criteria, I employed member checks. Member checks were done to ensure that the collected data are credible and, thus accurately reflect the interviewees' responses. In employing member checks, I provided each interviewee with a summary transcription of the interview a week after it was completed. After the interviewees reviewed the summary transcriptions, they were given the opportunity to correct errors within the summary transcription. They were also be given the opportunity to volunteer

additional information related to the topics addressed in the interview. Member checks helped to ensure that the data I collected are credible, i.e., accurately reflect respondents' views and perspectives.

In addition to member checks, I also employed methodological triangulation (Denzin, 1978). Methodological triangulation was achieved through the use of different types of data collection methods. Each method employed was used to uncover key information related to the study: "The use of different methods in concert compensates for their individual limitations and exploits their respective benefits" (Shenton, 2004, p. 65).

In addition to methodological triangulation, I also employed data triangulation. Data triangulation was performed in the data analysis phase of the study. In this phase, I compared data across data sources (document review and interview) to identify similarities or differences across the interview and document review data. This helped to confirm or negate the data collected across the multiple data methods.

Conformability. Establishing conformability in a study suggests that the results of study are free from researcher bias. Although this is not entirely achievable in research, researcher bias should be reduced as much as possible. To achieve conformability, I documented the research processes I undertook in the study step by step, particularly in the decisions I made with regards to the methods I choose and the methodological procedures I undertook. This type of technique for establishing conformability is referred to as an "audit trail." A data-oriented audit trail was used. This type of audit trail documents the ways in which data are gathered, analyzed, and reported in the study. It also demonstrates the processes which lead up to the formation of recommendations (Shenton, 2004, p. 72).

An additional technique that was employed to achieve conformability was reflexivity. Reflexivity requires a researcher to be explicit about his/her predispositions. Because a researcher's predispositions may bias the study, it is important that the researcher be explicit about his or her position in relation to the study. Employing reflectivity, I was aware that involvement in the evaluation in the role of evaluator may have biased the study. For example, the methods I chose to utilize or the ways in which I interpret the data may be affected by my perceptions of the study.

To reduce researcher bias, I addressed my role in the evaluation and was very explicit throughout the study about how my participation in the evaluation may have affected certain parts of the data analysis or interpretation of the study.

Limitations

Some limitations warrant consideration in this study. First, researcher bias in the study was possible given I was a part of the evaluation team for two of the three years of the evaluation. In this role, I assisted in data collection and analysis, as well as helped in writing the evaluation reports and presenting the evaluation findings to stakeholders (i.e., program staff and administrators). Due to my participation in the evaluation, I may have been biased in the data analysis and interpretation phase of the study; however, several steps have been taken to reduce this bias, including utilizing a rigorous research design; implementing multiple data collection methods; employing strategies for ensuring the collection of high quality data (i.e., the use of summary transcription, member checking). In addition, reflexibility will be pursed (discussed below).

Secondly, the timing of the study was also a potential bias to the study. The study is took place two years after the last evaluation was conducted. Because of the retrospective nature of

the study, the following steps were taken to ensure that respondents' answers reflected their true perceptions of the evaluation. Interviewees were provided with copies of the evaluation reports as well as a list of topics to be addressed in the interviews, and examples of dimensions of culturally responsive evaluation identified from the document review.

Thirdly, the use of secondary data, the three evaluation reports, may also be a potential limitation to this study. Secondary data have several limitations. For example, secondary data can be too general and therefore do not always provide the researcher with sufficient data for answering the research questions. In addition, secondary data can be flawed due to poor research designs or instruments which do not provide accurate or valid data. Therefore, secondary data analysis should proceed with caution. This research study compensates for some of the limitations of secondary data. First, the evaluation was a high-quality evaluation which utilized a rigorous evaluation design (i.e., consisting of multiple data collection methods). The rigor of the evaluation design, to some extent, ensures that the data are valid. Secondly, the evaluation was longitudinal, taking place over a three-year period. Evaluation data were collected on three separate occasions. The longitudinal nature of the evaluation, to some extent, ensured that the data are reliable.

CHAPTER 4

Evaluation Context

This chapter describes the evaluation context, which includes a description of the program and evaluation. The program is described in terms of its history, goals, components, activities, and program participants. In addition, the evaluation description includes the background, purpose, approach, key evaluation questions, audience, and criteria for judging program quality, design and methods. This chapter begins with an overview of the national context, including a review of STEM education, women and minorities in STEM fields, and an explanation for the lack of diversity present in STEM.

National Context

STEM education. National concern over the lack of qualified STEM workers in the United States has placed STEM education at the forefront of many national debates. Recent studies on the nation's economic competiveness have raised questions about the quality of STEM education. There is a widespread consensus that there is a need to improve the STEM education pipeline. A recent Educational Testing Service (ETS) survey reported that "61 percent of opinion leaders and 40 percent of the general public identify math, science, and technology skills as the most important ingredients in the nation's strategy to compete in the global economy" (United States Department of Labor, 2007, p. 2). Furthermore, the nation's lack of qualified STEM workers is a national problem. A recent report by the Department of Labor highlights the need for more qualified STEM workers and calls for the implementation of more effective strategies for the recruitment, retention, and support of students in the K-12 pipeline and beyond (United States Department of Labor, 2007).

The lack of qualified STEM workers and a rapidly growing concern over the nation's economic outlook has led different segments of society to take action. For example, the federal government has provided a significant amount of resources to improve the number of students and graduates in STEM fields and programs. Moreover, industries and firms have launched programs and/or initiatives which involve "curricular improvements, career-focused websites, mentoring programs, scholarships, and other incentives and supports" (United States Department of Labor, 2007, p. 6). Lastly, state governments have revamped state science standards to improve student achievement in science.[6]

Women and minorities in STEM fields. There is a widespread agreement that women and minorities are an untapped resource to science innovation and advancement in the United States. The incorporation of diversity in the current pool of STEM workers can increase the level of creativity, innovation and quality of scientific and technological products and services (Burke & Mattis, 2007). The importance of women and minorities in STEM occupations cannot be underestimated in today's economy; however, there are large gender and racial/ethnic disparities in the STEM workforce: "According to the U.S. Census Bureau's 2009 American Community Survey (ACS), almost three out of four jobs (72 percent) are held by non-Hispanic Whites, which is close to their overall representation in the U.S. workforce (68 percent)" (U.S. Department of Commerce, 2011). Additionally, while minority participation in STEM has increased in the past few decades, "progress has been marginal, neither steady enough nor substantial enough for the representation of minorities to approach parity with their presence in the U.S. population" (National Action Council for Minorities in Engineering Incorporation, 2008, p. 3).

[6] Several states have recently revamped their state science standards as a result of Race to the Top, a federal education program under the 2009 American Recovery and Reinvestment Act that provides money to states to implement education reform.

Recent reports highlight gender and racial disparities in STEM fields. While the representation of women in STEM has increased in recent years, participation has been limited to "low status" STEM fields: biological sciences, psychology, and social sciences. For example, from 1996 to 2006, women's percentage among science and engineering graduate students increased in degrees except computer sciences. Furthermore, women participation in high-status STEM fields has also declined. For example, between 1990 and 1998 bachelor's degrees awarded to women dropped in computer science. Further, the share of bachelor's degrees awarded to women in all STEM majors decreased from 30% in 1990 to 27% in 1998 (Clewell & Campell, 2002).

The participation of minorities in STEM reflects similar trends (Clewell & Anderson, 2002). While minority college enrollment has increased substantially over the past few decades, the share of bachelor's degrees awarded to minority students between 1997 and 2006 only increased slightly. For example, in 2006, 67.2% of Whites were awarded a bachelor's degree in comparison to 9.7% Asians, 8.7% African Americans, and 8.0% of Latinos (National Science Foundation, 2009). Moreover, the share of minority graduate students only increased by 2% from 9% in 1996 to 11% in 2006. Additionally, minorities' participation exceeded 20% only in psychology and the social sciences in either year (National Science Foundation, 2009).

Explanations for lack of diversity in STEM fields. The lack of diversity of women and minorities in STEM fields has been heavily studied (Blickenstaff, 2005). There is a general agreement among researchers that various factors have a negative impact on the participation of women and minorities in the STEM pipeline, and they include individual, school, home, and societal factors (Clewell & Anderson, 1991; Clewell, Anderson, & Thorpe, 1992, Clewell & Campbell, 2002; Oakes, 1990, & Oakes, J., Ormset, T., Bell, R., & Camp, P., 1990). Individual

factors refer to negative attitudes towards mathematics and science; low performance levels in mathematics and science courses and on standardized tests; limited exposure to extracurricular mathematics and science activities; and a lack of self-confidence and self-concept in math and science. School barriers refer to a lack of quality science instruction; failure to participate in advanced mathematics and science courses in high school; limited access to advanced courses; lack of information about and/or interest in mathematics-or science-related careers; teachers' stereotypes about math and science; and teachers' low levels of expectations. Home barriers refer to parents' lack of involvement and/or encouragement in math and science; parents' lack of exposure or familiarity with science and math; and parents' lack of understanding of the educational system. Societal barriers include gender stereotypes related to math and science, stereotype threat, and a lack of role models (Clewell & Anderson, 1991, Clewell, et al., 1992, Clewell & Campbell, 2002, Steele & Aronson, 1995, & Steele, 1997).

The larger national context above, particularly the lack of representation of women and minorities in STEM and the need to increase diversity within the STEM fields, were discussed in order to provide a context for this study, and particularly the G.A.M.E.S. summer camp. The discussion above illustrates the need and importance of the G.A.M.E.S. summer camp in the broader context of women and minorities in STEM. Additionally, it helps contextualize the design of the G.A.M.E.S. summer camp, providing a context for the programs' goals, components, and intended outcomes. Moreover, it situates the G.A.M.E.S. summer camp evaluation within the body of literature that addresses women and minorities in STEM.

Program Context

Description. G.A.M.E.S. was an annual week-long camp at UIUC designed to give academically talented middle-school-aged girls an opportunity to explore engineering and

scientific fields through demonstrations, classroom presentations, hands-on activities, and contacts with women in these technical fields. G.A.M.E.S. was sponsored by the following organizations: Abbott Laboratories, Alcoa, Ford Motor Company, John Deere, Motorola, and Shell, along with several STEM departments on the campus of the UIUC. In addition, G.A.M.E.S. was housed in the College of Engineering and organized by WIE program personnel. The WIE program also organize other programs for female and minorities in the STEM fields, focused on recruitment and retention.

History. G.A.M.E.S. was developed in 1995 by the Director of Special Programs in the College of Engineering. The purpose of the program was to expose talented young girls to engineering, and other STEM related fields. A secondary goal of the program was improve girls' motivation, interest, and academic abilities in STEM. At the time of the evaluation, the director of the program was an administrator in the College of Engineering whose participation greatly influenced the purpose, design and implementation of the program. The program director explained:

> As a civil engineer, I think I was more keen in rigor, I wanted a fun program but at the same time a program that would provide a strong foundation in the engineering basics, particularly given that we targeted gifted and talented girls.

In addition, the program director's personal background also impacted recruitment trends. G.A.M.E.S. was largely majority White in its first years of operation. As the program began to grow and the program director took a larger role in the program's recruitment efforts, the program began to diversify with larger numbers of minority participants being recruited into the program:

> As a Latina, I was very keen in increasing the minority participation of the group and I was able to achieve my goals. The year before I joined G.A.M.E.S. we had 60 girls and 15% minorities (8 girls), my last year in the program we had approximately 320 girls and 40% minority participation (almost 130 girls).

Goals: The goals of the program were multi-faceted. G.A.M.E.S. aimed to expose young girls to other young women pursuing careers in the sciences, as well as to give them the experience of living on a college campus. The program also sought to motivate young girls to pursue engineering and math majors and careers and gain confidence in their ability to succeed in such fields. According to the program director, G.A.M.E.S. had three main goals:

> To encourage girls' enthusiasm and interest in mathematics, science, and engineering; support the girls' sense of ability in these technical subjects; teach them that technical careers are accessible and interesting; present strong and compelling images of women engineers and scientists as intelligent and interesting; challenge the social isolation experienced by many academically talented young women by creating a sustainable community of interest around math and science; and encourage girls to consider a career in math, science or engineering; to increase the number of underrepresented minorities in the camp to reflect the representation in the State of Illinois and; to improve the retention rate of women engineering students at the College of Engineering at the University of Illinois. By hiring counselors and coordinators who are female college students in engineering, it provides these undergraduates with opportunities to become role models for each other and to encourage and support each other. The counselors and coordinators also become role models to the middle school girls, giving the college students an increased confidence in their abilities to complete their degrees.

Activities: Activities were offered through four programs exploring various STEM fields: a) Structures, b) Computer Science, c) Bioengineering/Chemical Engineering, and d) Bioimaging. In each program, campers worked on a team project, learned about various engineering, math, and science disciplines, participated in hands-on demonstrations, and met other young women interested in math and science. Each program is described below. Descriptions were drawn from evaluation reports, observations of camp activities, participants, and staff manuals.

Structures (*girls entering 6th or 7th grade in the fall*). The Structures Camp explored structural engineering and civil engineering, two different branches of engineering. Structural engineering is a special field branch of engineering that focuses on the design, construction and maintenance of structures society uses every day, from a family home to the tallest building in

the world, the Burj Dubai. Civil engineering involves designing, constructing, and maintaining the physical and natural world; this includes projects such as bridges, water, towers, roads, canals, buildings, and other structures. The Structures Camp provided campers with the opportunity to gain hands-on experience. The campers worked in small groups of eight to design and build a canoe. The campers were required to follow the engineering process of determining objectives, brainstorming, analysis, design, construction, testing, and evaluation to meet the project requirements. The campers were required to use green building construction, design a shelter that would allow four campers (girls) to sleep comfortably in it as well as stand in it; and design a shelter to withstand the Illinois weather elements. To guide campers in designing and building a canoe, they took classes in blueprinting, chemistry and materials, physics, and structures. At the end of the week, campers raced their canoes, and winners with the highest efficiency won.

Computer Science (*girls entering 7th or 8th grade in fall*). The Computer Science Camp explored the field of computer science, the study of theory, experimentation, and engineering that forms the basis of computers. The Computer Science camp also explored networks, binary numbers, edible circuits, and the engineering challenges of computer networks. The Computer Science Camp provided campers with the opportunity to develop computer graphics, interactive animations, and computer games. Campers participated in computer labs, hands-on programming, and building computer games using various computer programs such as *Squeak* and *Alice* software. For example, the campers learned how to place objects into the world (of the *Alice* program) and gained an understanding of algorithms, including how to write and implement one. Campers also participated in a final project in which they used basic computer concepts to create an animated storybook for preschoolers.

Bioengineering/Chemical Engineering (*girls entering 8th or 9th grade in fall*). The Bioengineering/Chemical Engineering Camp explored the fields of Bioengineering and Chemical Engineering, which team up to cure diseases, provide alternative sources of energy, and make an impact in developing countries. Campers applied both Bioengineering and Chemical Engineering to help prevent polio. To understand the many ways these fields approach this problem, the campers worked in the lab and completed activities throughout the week, which ranged from growing cells in cultures to building muscle models and experimenting with the movement of chicken legs. The campers learned that many systems in the body must work together to keep human beings alive and health every day. They also learned how engineering works to repair human muscles damaged by illnesses, such as polio, and the ways in which alternative sources of energy can be used in an economically and environmentally constructive manner to help prevent such diseases.

Bioimaging (*girls entering 8th or 9th grade in fall*). The Bioimaging Camp, which was held in collaboration with UIUC's Institute of Genomic Biology (http://www.igb.uiuc.edu/) used some of the world's best microscopes to investigate how small plant cells will respond to a big problem, climate change, and how that response affects us all. Campers were part of the ongoing effort to study global warming and food production by measuring the effects on plant cells of future levels of atmospheric carbon dioxide (CO2), the major greenhouse gas. Using the same advanced microscopy and instrumentation facilities that faculty and graduate students use at the Institute for Genomic Biology, campers imaged living plant leaves that have fluorescent mitochondria and chloroplasts and quantitatively analyzed them at the cellular level. Along the way, campers learned about optics, plant biology, chemistry, and image analysis (See Appendix G for a sample activity sheet)

In addition to group work, students also participated in a number of activities throughout the day. A typical camp day included: class sessions and hands-on demonstrations in the morning, team projects and activities, special guest lectures and engineering tours in the afternoon, and team projects and recreational activities such as African dancing, swimming, crafts, and other leisure activities in the evening. Further, campers were required to be in their room by 9:20pm and in bed at 10:00pm each night (See Appendix H for a copy of the daily camp schedule). Therefore, the camp schedule was very structured, with very little free time throughout the day.

Program Participants

Participants were recruited through various means. Brochures were emailed to the girls who had attended G.A.M.E.S. in the past and to those who had participated in academic enrichment programs around the state. WIE personnel also recruited at schools in the surrounding areas. Special attention was placed on recruiting students who were ethnically/racially and economically diverse. Scholarships were offered to those students whose participation might have been negatively impacted by program fees. Further, students were required to provide an interest in math and science through academic achievements and a strong commitment to developing their abilities in math and science. The application also required a reference letter from a teacher.

Each year, the program received more applications than open slots. Therefore, there were waiting lists of students who were interested in attending the camp. In 2007, two of the three camps had waiting lists. As a result of this increase in demand, the camp added an additional program in 2008, Bioimaging. The tables below demonstrate an increase in the overall student

participant in the camp. In addition, they demonstrate that Structures had the most participants each year while Bioengineering/Chemistry Engineering had the least.

Table 2

Percent of Participants by Camp in 2007

Program	Percent of Campers
Bioengineering/Chemical Engineering	16.0%
Computer Science	35.8%
Structures	48.2%
N	137

Table 3

Percent of Participants by Camp in 2008

Program	Percent of Campers
Bioengineering/Chemical Engineering	14.6%
Computer Science	25.9%
Structures	47.2%
Bioimaging	12.3%
N	215

Table 4

Percent of Participants by Camp in 2009

Program	Percent of Campers
Bioengineering/Chemical Engineering	13.7%
Computer Science	28.3%
Structures	40.4%
Bioimaging	17.6%
N	233

In addition to an increase in overall camp participation, there was also an increase in the number of minority participants in the camp, which was the result of the program director's efforts to recruit students from underrepresented backgrounds, namely African-American and Hispanic/Latino students. The tables below demonstrate the overall increase of minority participation over the years. From 2007 to 2009 there was an 11% increase in African-American participation and a 10.2% increase in Latino participation.

Table 5

Race/Ethnicity of Participants in 2007

Race/Ethnicity	Percent of Campers
African American	18.2%
Asian	20.4%
Caucasian	54.7%
Hispanic/Latina	2.2%
N	137

*Of the participants, 4.4% were mixed race.

Table 6

Race/Ethnicity of Participants in 2008

Race/Ethnicity	Percent of Campers
African American	20.0%
Asian	19.1%
Caucasian	33.5%
Hispanic/Latina	15.3%
N	215

*Of the participants, 0.5% were Native American and 11.6% did not provide their race/ethnicity.

Table 7

Race/Ethnicity of Participants in 2009

Race/Ethnicity	Percent of Campers
African American	29.2%
Asian	12.1%
Caucasian	37.8%
Hispanic/Latina	12.4%
N	233

*Of the campers, 8.6% did not provide their race/ethnicity

Evaluation Context

Background. The G.A.M.E.S. evaluation was not required by the College of Engineering, UIUC, or any outside agency. However, Women in Engineering (WIE) staff members requested the evaluation because they were interested in formally evaluating the program for the purposes of learning about the program, improving the program's design and

implementation, and disseminating information about the program to others in the engineering education community. The overall G.A.M.E.S. evaluation was intended to be comprehensive in scope—addressing issues of project design, context, implementation, and outcomes, as well as intentionally surfacing related value claims for reflection and critique.

Purpose. The primary purpose of the evaluation was to provide evaluative feedback to program administrators and staff regarding (a) the quality and appropriateness of program content, (b) the quality, effectiveness, and appropriateness of program pedagogy, and (c) the quality and meaningfulness of the program experience for its participants. A secondary purpose of the evaluation was to address issues related to the meaningfulness and benefits of the program for diverse kinds of children.[7] This feedback was intended to enhance stakeholders' understanding of the program and to provide information useful for program improvement. The program director described the purpose of the evaluation in the following way.

> I wanted to understand if the camp goals were being met and to the determine areas of improvement that would have brought us closer to our goals. During my tenure, the evaluation focused on programmatic areas of the camp itself, and also on the participants' success. The evaluators agreed that the purpose of the evaluation was to improve the program; however the purpose of the evaluation was also to document the program success.

Evaluator comments on the evaluation's purpose are below:

> We did the evaluation for three years. I think the first year it was really formative. The program director had never evaluated the program before and she felt that it was a really good program, but I think that at first she wanted an evaluation that would tell her what was working and what needed to be improved. Then the purpose of the evaluation shifted because she wanted to expand the program, and she wanted to get more funding, and then she wanted to know and be able to show the impact of the program on the girls' interest in science.

[7] The two purposes (primary and secondary) of the evaluation were developed using an Educative, Values-Engaged Approach to Evaluating STEM Education Program. More information can be found in an article by Jennifer C. Greene, Lizanne DeStefano, Holli Burgon, and Jori Hall, "An Educative, Values-Engaged Approach to Evaluating STEM Educational Program" in the New Directions for Evaluation Journal no. 109, Spring 2006. This approach is now referred to as A Values-Engaged, Educative approach. However, in this dissertation it will be referenced by its original name.

> From my perspective, the G.A.M.E.S. program was very successful and they wanted an official evaluation to get grants and money. So they were interested in documenting their success. They wanted an outside person to go in and evaluate the effectiveness of the program in an official report.

Evaluation Approach. The overall evaluation of the G.A.M.E.S. program utilized an Educative, Values-Engaged Evaluation approach to evaluation (Greene, DeStefano, Burgon, & Hall, 2006). The EVEN approach defines a high-quality STEM education program at the intersection of content, pedagogy, and equity and is inclusive of multiple perspectives and interests in STEM education. That is, high-quality STEM education programs address cutting-edge, critical scientific context, implement effective pedagogy, and successfully meet the needs of diverse learners. The EVEN approach was selected because its goals were well matched with those of the G.A.M.E.S. summer camp. Namely, G.A.M.E.S. aimed to provide: 1) high quality instruction in math and science; 2) high quality pedagogy that is appropriate with the backgrounds of the participants; and 3) meaningful experiences and benefits in math and science (i.e. promote positive attitudes toward math and science). The EVEN approach was also selected because it provided the opportunity to further study the EVEN approach to evaluation. The EVEN approach was developed and tested through a series of NSF grants. The G.A.M.E.S. evaluation provided an additional opportunity to further field test the EVEN approach.

While the primary evaluation approach was EVEN, the evaluation also employed culturally responsive evaluation. Namely, the evaluation described and explained the cultural content of the program (Frierson, Hood, & Hughes, 2002; Thomas, 2004). Further, the evaluation was responsive to culture and context, two key dimensions of cultural responsive evaluation, through the use of various culturally responsive evaluation strategies. The evaluation employed a culturally diverse evaluation team, robust evaluation design of qualitative and

quantitative methods, and culturally sensitive data collection instruments. In addition, the evaluation engaged stakeholders, analyzed data from a culturally responsive perspective, and distributed the evaluation results widely. While the primary evluation approach was not culturally responsive evaluation, clear dimensions of culturally responsive evaluation were present throughout the evaluation from its initial conceptualization to its implementation.

Key Evaluation Questions. The evaluation addressed four questions which were developed based on the purpose of the evaluation, which was to improve the program design. In addition, the evaluation questions were developed from the EVEN approach to evaluation and key dimensions of culturally responsive evaluation. The evaluation questions included:

1. What is the quality of the content covered during the G.A.M.E.S. summer camp, and is this content appropriate for program participants (i.e., to what degree does the content of the program match the program participants' developmental and learning profiles)?
2. What is the quality of the pedagogy (instructional design) utilized during the G.A.M.E.S. summer camp; is this pedagogy effective (i.e., are students engaging in active learning); and is this pedagogy appropriate for program participants (i.e., to what degree does the pedagogy of the program match program participants' developmental and learning profiles)?
3. In what ways and to what extent does camp participation influence participants' attitudes toward engineering and STEM more generally, as perceived by teachers, parents, and the students themselves?
4. In what ways and to what extent does camp participation effectively support camp counselors as they matriculate through the UIUC engineering degree program?

The evaluation questions were designed to be culturally responsive in nature. Therefore, components of culture responsive evaluation were incorporated into each evaluation question. For example, the first evaluation questions addressed the level of quality and appropriateness of the program content based on the diverse learning and developmental backgrounds of the participants. Further, the second evaluation question addressed the quality and effectiveness of the program's pedagogical methods, again based on the diverse learning and developmental backgrounds of the participants. Moreover, the third evaluation question addressed the participants' attitudes toward STEM from various perspectives (i.e. as perceived by teachers, parents, and the student themselves). Lastly, the fourth evaluation question addressed the impact of camp participation on the matriculation of the camp counselors.

Of the four evaluation questions, the first and second evaluation questions were the most culturally responsive in that they addressed the diverse backgrounds of the participants. For example, evaluation questions one and two addressed the different developmental and learning profiles of the participants. In addition, they also addressed differences in program impact based on the different developmental and learning profiles of the participants. Further, all four evaluation questions addressed the diversity of program participants beyond their different developmental and learning profiles. The evaluation questions addressed the racial and ethnic backgrounds of the program participants, namely race and socio-economic diversity. In addition, language, hometown, and school diversity were also addressed in the development of the evaluation questions. In reference to the evaluation questions one evaluator stated:

> We looked to see if there was a difference in the outcomes of the different racial groups, and I think it was important to see how to judge the effectiveness of the program on the different groups to try to address is there was a difference in the participants' experiences in the program.

Audience. The primary audience for the evaluation was the client of the evaluation. The client of the evaluation was also the program director. The program director was the person who requested the evaluation, and she was also the contact person for the evaluation. The evaluators were in direct contact with the program director throughout the evaluation process to ensure that her requests for the evaluation were met and to include her in the evaluation decisions, which is a key dimension of culturally responsive evaluation, to include stakeholders in the evaluation.

In addition to the client of the evaluation, the evaluation team also designed the evaluation to target other evaluation audiences. Evaluation audiences considered in the evaluation in addition to the program itself included camp counselors, program administrators, participants, and parents. More specifically, camp counselors were female undergraduate students pursuing STEM degrees at UIUC. Administrators were employees of the College of Engineering or were employed in other departments on the UIUC campus who played a role in implementing the program (i.e. facilitated staff trainings, hired program staff, organized program activities, etc). Program participants and parents included the middle school students who participated in the program and their parents. Therefore, the evaluation audiences ranged widely and included all individuals who participated in the program who could have benefitted from the evaluation results.

Criteria for Judging Program Quality

The evaluation employed an EVEN approach criteria for judging program quality. This criteria requires that STEM education programs are judged from four vantage points: (a) the quality of the project content and instructional design for learners in the contexts served; (b) the contextual relevance and power of the project design; (c) the advancement of the interests of under-represented and under-served groups; and (d) the quality of project implementation and

outcomes. Similar to the evaluation questions, the criteria for judging program quality also incorporated key dimensions of culturally responsive evaluation. Of the four vantage points, "the advancement of the interests of under-represented and under-served groups" most closely represented the goals of culturally responsive evaluation.

In addition to the four vantage points, an EVEN approach criteria for judging program quality employs specific criteria. These criteria were related to the quality of the context and pedagogy of the program and experiences of the participants in the program. The criteria included:

1. The program provides high quality instruction that is appropriately connected to the developmental and learning profiles of participants.
2. Program content is of high quality, developmentally appropriate, and congruent with the goal of promoting student achievement in advanced mathematics courses.
3. The program provides meaningful learning experiences for participants.
4. The program increases equity by recruiting children from underrepresented groups with access and opportunity.
5. Participants demonstrate meaningful benefits from their experience with the program (e.g., improved mathematics self-efficacy, improved attitude or learning habits, social benefits).

Of the criteria specific above, the two most closely aligned with culturally responsive evaluation included: a) the program increases equity by recruiting children from underrepresented groups with access and opportunity, and b) the program provides meaningful learning experiences for participants. The first criteria addressed the cultural diversity of the

participants with a goal of advancing the interest of underrepresented participants, and the second addressed the various experiences of the participants in the program.

In addition to the criteria above, the content and pedagogy of the program was compared to the national standards included in the: 1) Standards for Technological Literacy: Content for the Study of Technology and 2) National Science Education Standards (NCES), respectively. Both standards address issues as related to student progress in technology and science. The Standards for Technological Literacy outlines what students should know, understand, and be able to accomplish in science technology in Grades K-12. The Standards for Technological Literacy offers standards and criteria for judging student progress in five areas of technology and are organized into the following categories: 1) The Nature of Technology, 2) Technology and Society, 3) Design, 4) Abilities for a Technological World, and 5) The Designed World (ITEA, 2007). To demonstrate, the Structures Camp met the *ITEA Content Standard #1: Students will develop and understanding of attributes or engineering design*. In the Structures, activities met the criteria below:

1. Students will learn how to use the scale and other mapping skills.
2. Students will understand that ratios are used to create scale models and/or blueprints.
3. Students will describe what can be done to enhance the chances a cardboard boat will not disintegrate in the water.
4. Students will use physics principles to help make the boat.
5. Students will get practice creating a model boat.

Recommendations are also provided within the ITEA Content Standards. Embodied within the recommendations were issues of equity, which added to the cultural responsiveness of the evaluation. One of the recommendations was "teachers and curriculum developers should

address minority, gender, and equity issues to ensure that students are encouraged and motivated to succeed in the study of technology" (Technological Literacy Standards, p. 18).

The National Science Education Standards (NCES) also outlines what students should know, understand, and be able to accomplish in scientific literacy in Grades K-12 (NSES, 1996). The NCES standards for teaching "provide criteria for making judgments about progress toward the vision; they describe what teachers of science at all grade levels should understand and be able to do" (p. 27). The content and pedagogy of the G.A.M.E.S. summer camp were compared to the teaching standards outlined in the National Science Education Standards. For example, Teaching Standard A: Teachers of science plan an inquiry-based science program for their students and the four camps requires teachers to: 1) develop a framework of yearlong and short-term goals for students; 2) select science content and adapt and design curricula to meet the interests, knowledge, understanding, abilities, and experiences of students; 3) select teaching and assessment strategies that support the development of student understanding and nurture a community of science learners; and 4) work together as colleagues within and across disciplines and grade levels (NSES, p. 30). Below is an example of the alignment between a teaching standard and the different camps.

Table 8

Comparison of the G.A.M.E.S. Camps and NSES

NSES: Plan an inquiry-based science program for their students.	
Bio/Chemical Engineering	After several lessons in topics such as cell culture, muscle repair, actin, biomechanics, filtering, bioreactors, distillation, and polymers, students are asked to develop a plan to manufacture and distribute a new polio vaccine to the Dominican Republic.
Structures	After lessons in physics, materials, engineering design, and blueprints, students work in groups to develop a blueprint and build a cardboard canoe to race in a pool using various engineering concepts learned.

(Table 8 continued)

Computer Science	After watching demonstrations of writing code to make computer generated objects move and make sounds, students work in pairs to develop a video game or multimedia story using the Squeak/Alice program.
Bioimaging	After several lectures in topics such as plant biology, global warming, the effect of CO_2 on crops, the fundamentals of optics, making measurements on digital images, collecting tissue samples from plants, and communications skills, students are asked to develop a design for crops to meet the nutritional needs of the population on a deserted island.

The NCES also embodies issues of equity and are highlighted in the purpose of the standards. "The intent of the *Standards* can be expressed in a single phrase: Science standards for all students. The phrase embodies both excellence and equity. The Standards apply to all students, regardless of age, gender, cultural or ethnic background, disabilities, aspirations, or interest and motivation in science" (NCES, p. 2) Again, the NCES standards demonstrated dimensions of culturally responsive evaluation, which not only made the NCES an appropriate criteria for judging program quality, but also increased the evaluation's focus on culture.

Evaluation Design and Methods

The evaluation employed a mixed methods design. Mixed methods designs are used to achieve various purposes (i.e. complementarily and expansive) (Greene, 2007). In this evaluation, mixed methods were use to (a) generate a comprehensive, in-depth understanding of the program of interest; (b) produce a fair representation of the complexity of a program; and (c) tap into various aspects of the same phenomena (Greene, 2007). The multiple methods employed in the evaluation included: a) document review, b) observations, c) surveys, d) focus groups, and e) interviews. These methods rendered both qualitative and quantitative data.

Document review. Lesson plans and project manuals for the campers were collected and analyzed for descriptive information about program design, recruitment, participant selection, and key characteristics of program participants.

Observations. Lesson observations were conducted to capture a descriptive record of the structure, content, and activities of the program, and the character of the students' participation in them. For observations of students' participation, 2–3 program participants were randomly selected for each session for close observation, so that a large sample of participants would have been observed after multiple observations. For all observations, a detailed description of participants' engagement was produced (See Appendix C for Observation Guide).

Surveys. All program participants completed brief pre- and post-surveys. The surveys focused on participants' perceptions about, and interest in, math and science. The survey also assessed participants' content knowledge in the subject matter pre-and post-camp participation. Each questionnaire required 30 minutes to complete (Please see Appendix I for Participant Survey).

Focus groups. The purpose of the focus groups was to gather data on the participants' experiences with the G.A.M.E.S. summer camp as they related to program content and pedagogy, contextual characteristics relevant to program implementation, and the impact of the program on student performances. The evaluation team conducted two focus groups of randomly selected program participants. In addition, separate focus groups were conducted for the various race/ethnic groups and camps. Focus groups required 30 minutes to complete. The size of the focus groups ranged from 5 to 12. Please see Appendix J for Focus Group Interview Protocol.

Interviews. Parents and camp counselors were interviewed in person or via the phone to gather information on perceptions of the quality and benefits of the G.A.M.E.S. summer camp

participation. The evaluators selected both a random and purposeful sample of parents to interview. Separate interviews were conducted with parents for the various race/ethnic groups. Interviews with the Latino parents were conducted in Spanish. In addition, a random sample of camp counselors from each program were interviewed. The interviews were brief, requiring 10–15 minutes to complete (See Appendix K and L for Parent Interview Protocol and Camp Counselor Interview Protocol, respectively).

Table 1 on Page 102 presents the linkages between key evaluation questions, the information required from the evaluation questions, and the methods utilized for gathering data. This matrix was developed by the evaluators and was shared with the program director. The purpose of the matrix was to ensure that both the evaluators and program staff agreed on the evaluation questions and methods for data collection

Table 9

Key Evaluation Questions and Methodology

Key Evaluation Question	Information Required	Method	Sample
Q1. What is the quality of the content covered during the G.A.M.E.S. Summer Camp, and is this content appropriate for program participants?	Instructional Content Content & standards alignement Participant experience of content	Document analysis and observations Interviews and focus groups of program counselors and participants	All lesson plans and lab manuals. Sample of 13 camp counselors/coordinators and 18 campers.
Q2. What is the quality of the pedagogy utilized during the G.A.M.E.S. Summer Camp, is this pedagogy effective, and is this pedagogy appropriate for program participants?	Instructional Design Quality science pedagogy Participant experience of activities	Document analysis, literature review Interview camp counselors and staff Focus Groups	All camp lesson plans Sample of 13 camp counselors 18 campers
Q3. In what ways and to what extent does camp participation influence participants' attitudes toward engineering and STEM more generally, as perceived by teachers, parents, and the students themselves?	Participants' knowledge of study skills, career planning and college application procedures	Survey student participants Interview parents Conduct focus groups	138 student participants Random sample of at least 25% of parents
Q4. In what ways and to what extent does the camp participation effectively support camp counselors as they matriculate through the UIUC engineering degree program?	Counselors' academic and career planning behaviors, attitudes, and achievements	Counselor interview Focus Group	Sample of 13 counselors

CHAPTER 5

Findings: Preparing for, Designing, and Implementing Culturally Responsive Evaluation

This chapter provides a detailed description of the findings of the case study. The main research question in the case study included: *How do evaluators plan, design, and implement culturally responsive evaluations?* To answer the primary research question, I examined how the G.A.M.E.S. evaluation employed strategies for conducting culturally responsive evaluation. More specifically, I compared the evaluation strategies used in the evaluation to the strategies outlined in the 2002 and 2010 User-Friendly Handbooks for Project Evaluation (NSF, 2002; 2010). The following culturally responsive evaluation strategies were compared:

1. Assemble a culturally diverse evaluation team
2. Identify stakeholders
3. Engage stakeholders
4. Identify the purpose(s) and intent of the evaluation
5. Collect background information on the evaluand
6. Frame the right questions
7. Select an appropriate evaluation design
8. Analyze the data from a cultural lens
9. Disseminate and use the results
10. Ethical Considerations

Assemble a Culturally Diverse Evaluation Team

The creation of a culturally diverse evaluation team is critical to culturally responsive evaluation. Evaluations of programs that serve ethnically and/or racially diverse students call for the creation of a multi-ethnic evaluation team (Frierson et al., 2002). In culturally responsive

evaluation, evaluators must have a shared lived experience in order to "truly" understand the cultural context of the program and its participants. Evaluators who have shared lived experience (i.e. share the same racial/ethnic background to the participants) are more likely to be aware of, respond to, and accurately interpret cultural nuisances (i.e. communication and relationships styles of various cultural groups). Moreover, evaluators who have a shared lived experience are more likely to understand the cultural context in which the program operates, including the cultural experiences, norms, and traditions of the participants (Thomas, 2005; Frierson et al., 2002; 2010; Prado, 2011).

The G.A.M.E.S. evaluation met this expectation of responsive evaluation. The G.A.M.E.S. evaluation employed a culturally diverse evaluation team each year was completed. The evaluators who participated in the evaluation not only represented the racial/ethnic backgrounds of the participants, but they also had previous experience in program evaluation in diverse evaluation contexts. The racial/ethnic backgrounds of the evaluators included: African-American, Latino, and Caucasian. In addition to the racial/ethnic backgrounds, the evaluators also possessed additional skills which were beneficial for data collection. For example, some evaluators were bilingual. All the evaluators agreed that the evaluation team was culturally diverse:

For all years of the evaluation the evaluation team was pretty diverse. The first year there were six evaluators, half were African American and half were Caucasian. The next two years we had African American and Hispanic evaluators, which was good, because we started to get more African American and Hispanic participants.

> I think that the evaluation team was diverse. I brought people into the evaluation that I felt represented the different cultural groups involved in the camp.

The program director agreed with the evaluators that the evaluation team was culturally diverse, and represented the various racial/ethnic backgrounds of the participants. One evaluator stated" "The evaluation's team diversity reflected the diversity of the participants." Another evaluator stated: "We used diverse evaluators or diverse data collectors so that that could connect and relate to and understand the different populations that were involved in the evaluation."

In accordance with a culturally responsive evaluation approach, the G.A.M.E.S. evaluation assembled a culturally diverse evaluation team to ensure the evaluators had an understanding of the culture of the program, including its participants, and to increase the likelihood the evaluators could relate to and identify with the cultural nuances of the participants, for example, non-verbal and verbal behaviors.

Identify Stakeholders

The identification of stakeholders is an important step in the design of an evaluation. In culturally responsive evaluation, the relevant interests and concerns of the program stakeholders are taken into account in the evaluation. Because some groups of individuals are more likely to be represented within the evaluation more than others, for example people with power and privilege, is it the responsibility of the culturally responsive evaluator to identify and take into account all stakeholders in the evaluation who have a stake in the evaluation. Failure to take into account some stakeholder groups "might result in failing to capture critical contextual aspects of the project under study, which potentially lead to inaccurate judgments and conclusions" (Frierson et al, 2010, p. 81).

In the G.A.M.E.S. evaluation, the identification of stakeholders occurred in the early phases of the evaluation. The primary stakeholder of the evaluation was the client of the evaluation, who was also the director of the program. She worked in collaboration with the

evaluation team to design the evaluation. For example, she provided sound advice to the evaluation team on different aspects of the evaluation, including the framing of evaluation questions and dissemination of the evaluation results. In this evaluation, evaluation team members ensured that program stakeholders were included throughout the evaluation process (i.e. participants, parents, administrators, staff, etc.) One evaluator stated:

> We described stakeholders broadly, I mean part of the stakeholders were program staff, parents, the participants themselves, and other faculty and engineering administrators that kind of sponsored the program. We tried to engage them as much as possible.

Stakeholders who played a more subtle role in the evaluation included program staff other than the program director, as well as program participants and parents. Although these individuals did not play a substantial role within the evaluation, their interests were always considered by the evaluation team and incorporated into the evaluation. Further, additional stakeholders included in the evaluation were faculty and administrators from the College of Engineering and the engineering community at large (i.e. researchers, scholars, etc.).

The program director was not only the primary evaluation stakeholder, but her role in the evaluation was crucial. She provided feedback on the design and implementation of the evaluation. Her interest in the evaluation was also well-known, and she developed a good working relationship with the evaluation team over time, which created a positive and welcoming environment for the evaluation to take place in. Evaluators provided the following comments regarding the program director.

> I had a good personal relationship with the program director so I think pretty much the evaluation was pretty positive thing.

> The program director [and one of the evaluators] seemed to be peers and the program director seemed to have great for what [one of the evaluators] had to say.

> The program director was the person that wanted the evaluation and I think the dean was very supportive of the program because the program I think was kind of a shinning jewel.

Engage Stakeholders

An important component of culturally responsive evaluation is stakeholder engagement. In culturally responsive evaluation, evaluators engage stakeholders in throughout the evaluation, from the formation of the evaluation questions to the dissemination of the evaluation results (Frierson, et al., 2011). When seeking stakeholder engagement in culturally responsive evaluation, special attention is given to communities and/or groups of individuals who have traditionally been underrepresented and/or misrepresented in evaluation (i.e. racial/ethnic minority groups, the poor, etc.) (Frierson et al, 2002; Thomas, et al., 2004; Ryan, et al., 2007; Frierson, et al., 2010; Samuels & Ryan; 2011).

In the G.A.M.E.S. evaluation, stakeholder involvement was an important component of the evaluation. Evaluation stakeholders were involved in every phase of the evaluation process. Prior to the evaluation, program staff met with the evaluation team to plan for and design the evaluation. For example, the program staff met with the evaluators on several occasions to identify the purpose of the evaluation and the evaluation questions. Evaluators commented in relation to the evaluation's purpose and questions:

> We talked to the program director to see what kind of information she was interested in getting. We also went over the questions ahead of time with her.
>
> In this case, we did engage the staff in coming up with the questions, and I think the [program director] had a very clear sense of what she wanted and it was really around issues of diversity.

In subsequent meetings with the evaluators, the program staff also gave input on other aspects of the evaluation, which included data collection methods and instruments (LaPoint & Jackson, 2004). In addition, the types of evidence that would be considered credible in the evaluation were also discussed in meetings between the evaluation team members and program

stakeholders. In addition, discussions of how the evaluation results would be used and disseminated also took place. One evaluator provided the following comment:

> They were engaged in figuring out what evidence would be compelling to them. I think that they also gave input to the instruments and how we were going to actually collect the data.

During the evaluation, program stakeholders were also equally engaged throughout the evaluation. The program director played several roles during the evaluation. For example, she helped ensure that the evaluation was running smoothly, and that the evaluators had access to various materials and classrooms. The program director also assisted in the data collection process by offering suggestions to the evaluators on how data could be most efficiently collected (i.e. to conduct the focus groups after the lunch period in order to achieve higher participant rates). Further, the program director provided feedback to the evaluators on which lesson plans and activities to observe that were not only of interest to her but also relevant to the evaluation's goals. Further, she also ensured that the concerns and interests of the other program staff, including the participants, were adequately addressed by the evaluators by informally meeting with them throughout the evaluation (i.e. through checking in).

After the evaluation was complete, program stakeholders were also included in the evaluation. At the conclusion of the program, the evaluation team met with the program director to reflect on the evaluation and to address any issues which needed immediate attention. Further, after the data analysis was complete, the evaluation team met again with the program director to provide a preliminary analysis of the data. After the evaluation report was complete, the evaluation met a final time with the program director to present the evaluation findings. At the meetings, the evaluators gave the program director the evaluation report, and the program director had the opportunity to ask questions or make comments on the evaluation report,

including the results and recommendations. The evaluators also gave the program director an opportunity to request further data analysis; on several occasions the program director requested the evaluation data to be analyzed differently, for example by subgroups (i.e. camp). The program director commented on the ways in which the evaluators involved her in the evaluation:

> I met with the evaluation team after the evaluation was complete, and everything was fine. Then the process they followed was they would send me the evaluation report for me to review, and after I looked at it they would invite me to a meeting…And something that I like about the whole thing is that they would go through the findings of the evaluation, and based on the findings, they would make recommendations.

At the conclusion of the evaluation, the evaluators also involved other program stakeholders in the evaluation, in addition to the program director. These included program staff (i.e. camp counselors), faculty, and administrators from the College of Engineering. In addition to the evaluation report, the evaluation team also completed formal presentations of the evaluation to the program director, and the other evaluation stakeholders listed above. During the evaluation presentations, all program stakeholders were given the opportunity to provide comments or feedback on the evaluation. The evaluation team also responded to questions from the audience. In some cases, program stakeholders provided suggestions to improve the evaluation in the future.

Overall, the evaluation was designed from the perspective that stakeholder involvement was critical to the evaluation and evaluation use. Thus, the program stakeholders were involved in every aspect of the evaluation. Evaluators commented on stakeholder engagement and provided the following quotations:

> I think that the thing that kind of always drives me the most when designing an evaluation is use so what can we do to make the evaluation useful and so in this case [the G.A.M.E.S evaluation] it did involve really engaging the staff in the evaluation.

> I felt like we involved [the program director] in every aspect of the evaluation. I mean, she looked over everything and made comments, and we did consider her opinion and

> take her advice a lot in the evaluation. I cannot say she came up with some of the questions, but I felt like when I was writing it, I trying to address some of things the she was talking about.

Identify the Purpose(s) and Intent of the Evaluation

Identifying the purpose and intent of the evaluation is important in any evaluation. Program evaluation seeks to address two primary questions: (1) is the project being conducted as planned and is progress being made toward meetings its goals? and ultimately, (2) how successful is the project it reaching its goals? (Frierson et al, 2010, p.83). In program evaluation, there are different types of evaluations that can be designed to provide answers to the aforementioned evaluation questions and they include process, progress, and summative evaluations (Frierson et al., 2002; Frierson, et al., 2010). Moreover, in culturally responsive evaluation, the cultural context of the program is taken into account to examine the relationships between the project activities (in process evaluations), to determine whether or not the project is meeting its goals and objectives (progress evaluations), and to examine program impact on the participants (summative evaluations). Finally, the culture of the program and participants are included in every type of evaluation: process, progress, progress, and summative.

As mentioned earlier, the purpose and intent of G.A.M.E.S. evaluation was identified through meetings with the program director. These meetings explored the client's interest and needs for the evaluation. The purpose of the evaluation was established early in the process and in collaboration with the program director. The purpose of the evaluation was both formative and summative in nature (LaPoint & Jackon, 2004). The first evaluation was formative because it provided feedback to the program director on the design and implementation of the program activities. Formative evaluation also assessed the strengths and weaknesses of the program. The second and third evaluations were summative because they provided feedback to the program

director on program outcomes. Summative evaluation also assessed the extent to which the program was meeting its goals and objectives. One evaluator commented:

> We did the evaluation for three years; I think the first year it was really informative. The director of the program had never done an evaluation, and she felt like it was a really good program. At first, she wanted an evaluation that would tell her what was going well and what needed to be improved, so I think initially it was informative. And I think that then it shifted because she wanted to expand the program, and she wanted to get more funding. And then she really wanted to show the impact of the program on girls interested in science, and then also on any kind of intercultural relationships, and so then I think it became more toward a more summative evaluation to look at those two things interest in science and the intercultural piece.

Feedback from both formative and summative evaluation was discussed with the program director at various stages of the evaluation process. Formative feedback was provided to the program director during the evaluation. The evaluation team met with the program director informally and formally to discuss program activities and/or participants' experiences as observed throughout the evaluation. Additionally, as mentioned above, the evaluators discussed with the program director issues in which needed to be addressed immediately. Formative feedback was also provided to the program director at the end of the evaluation. This feedback was provided in meetings as well as presented in the evaluation reports and presentations given to the evaluation stakeholders (i.e. program staff, faculty and administrators in the College of Engineering).

The evaluation team provided summative feedback to the program director at the end of each evaluation. This feedback was related to the program's impact on participant outcomes, including the participants' attitudes toward STEM. Evaluation reports were provided to the program director, which included a list of evaluation results and recommendations for program improvement. This information was also presented to stakeholders through formal presentations given at the College of Engineering. Feedback provided to stakeholders throughout the

evaluation help to ensure the evaluation's purpose was clearly articulated, and the goals of the evaluation were being met.

Collect Background Information on the Evaluand

Prior to the implementation of a culturally responsive evaluation, evaluators collect background information on the evaluand (Frierson et al., 2002; Frierson, et al., 2010). The evaluand in an evaluation can include the culture of the program and its participants. Data on the evaluand is collected to acquire a full understanding of the cultural context of the program. This type of data collection requires that evaluators spend a significant amount of time in the evaluation context, so that they can get acquainted and build relationships with program stakeholders. In addition, culturally responsive evaluators seek to develop respect for program staff and participants in the evaluation context. In addition, data collection prior to the evaluation enables evaluators to develop familiarity with the cultural norms, traditions, and beliefs of the participants (Frierson et al., 2002; Frierson, et al., 2010).

In the G.A.M.E.S. evaluation, information on the program was collected throughout the evaluation process. First, background information on the program and participants was collected through informal and formal interviews with program stakeholders prior to the evaluation. Both informal and formal meetings with program staff served to give evaluators an understanding of the program context: political, social, economic, and historical dynamics of the program. In addition to interviews, background information on the program and participants was also gathered through meetings. Meetings with the program staff revealed important information on the goals of the program, the different program activities, and participant demographics.

In addition to interviews and meetings, information on the program and participants was also gathered through document review. Program materials were reviewed and analyzed by the

evaluators to capture a descriptive record on the background of the program and participants. In the document review, a variety of program materials were reviewed: program website, program manuals, and promotional materials (i.e. brochures). Further, document review of program materials revealed important information about the different programs (i.e. Structures, Computer Science, etc), camp schedules, and activities. Additionally, promotional products such as the programs website and brochures rendered information on the application procedures, deadlines and eligibility. This information was used by the evaluators to develop a complete picture of the program with an additional focus on its culture.

Information on the program was also collected during the evaluation. This information was collected in multiple venues (Frierson et al., 2002; Frierson, et al., 2010) and was related to the program's geographic location, its history, and its political and economic context. For example, throughout the evaluation, the evaluators met with the program staff at the College of Engineering who sponsors and organizes the program yearly. The evaluators met with the program staff in the College of Engineering as opposed to somewhere else on campus in order to become familiar with not only the culture of the department and college, but also the interpersonal relationships in this cultural context. One evaluator commented:

> We tried to get a really good understanding of what the culture of the college, the culture of the program, and the culture of the people within the program, so it was very much embedded; there was a lot of dialogue.

In addition to meetings in the Department of Engineering, evaluators conducted informal interviews with individuals during the evaluation. They were conducted with program staff, administrators, faculty, parents, and participants. The purpose of these interviews was to gather information on the cultural context of the program from multiple stakeholder groups. The informative rendered provided the evaluators with data on the program and its culture from

various perspectives, including information from individuals who were not necessarily in leadership positions (Frierson et al., 2010, p. 79).

In addition to informal interviews, evaluators also conducted first hand observations of the activities of the program. The purpose of these observations was to capture a descriptive record of the culture of the program, with a focus on the camp activities. Evaluators observed various camp activities, including staff training and meetings, lessons plans, program activities, camp orientations, award ceremonies and student presentations. The program director reported:

> The evaluators would come to our meetings for training. They would come in and explain to the girls what they were trying to do and what they wanted to do. They would also explain to the girls why the evaluation was important and why they were doing it. Any time I had a question I would ask them [the evaluators]. I also remember the counselors, whenever they found issues that they thought were interesting, they would go and talk to the [evaluators], or if somebody had a comment about the program they would go and talk to the [evaluators], so in many ways the [evaluators] did get input from the girls and the counselors, but it happened more informally.

In addition, the evaluators reported that they paid attention to the cultural context of the program in the informal observations they completed. Further, this context was addressed in the informal interviews they administered with the parents (as mentioned above). For example, one evaluator commented:

> We tried to have an understanding of the cultural context of the program. We talked about it a lot; we brought it to the forefront of the evaluation. We were not just saying do they have a greater interest in STEM, but we looked at it, and when I was watching the final presentations of the kids, I looked at the color of the people in the audience: I looked at where they were sitting; I look at how they were responding; you know, I talked to them in the lobby, so we brought it [culture] to the forefront as something that was of interest to us.

In addition to observations of camp activities, evaluators completed session observations of camp lessons. These observations were completed to render information on the participants' engagement with the activities. In addition, the session observations provided information on the culture of the participants in that the evaluators observed the participants' behaviors, attitudes,

and beliefs. More specifically, the evaluators observed how the participants communicated with one another and behaved toward each other in the context of the program and through the session observations. One evaluator stated:

> We spent a lot of time on-site collecting information on the culture of the camp, but also the culture of participants. We looked at how they communicated, how they interpreted certain behaviors, you know, body language and the language of the students.

In addition to formal evaluation activities such as observations and interviews, informal conversations with the program director also helped the evaluators to gain a broader understanding of the cultural context of the program and its participants. During informal conversations with the evaluators, the program director often spoke about her personal experience as a women and minority in STEM. Moreover, she also spoke to the evaluators about the need for programs like G.A.M.E.S. Evaluators commented on the background of the program director and its role on their understanding of the program context.

> I talk to the program director. We had a really good rapport, and she often talked about what it was like to be a woman and to be a minority in the engineering field, so based on that that and some of things that I read about helping maintain women and minorities in engineering, kind of gave me some background on the cultural context of the program.

> I remember one of the evaluators said that the Latina girls liked to do things with their hands and that would have not been an insight that I would have had. That is something the [Latina evaluator] brought, and had she not have been on the team, we [the evaluation] team might have not had. The program director who was also Latina also brought that to the context.

Frame the Right Questions

Framing the right evaluation questions is an important part of a successful evaluation. In culturally responsive evaluation, the evaluation questions are important to the stakeholders. That is, the evaluation questions address issues relevant to the stakeholders' interest in the evaluation. Moreover, in culturally responsive evaluation, the evaluation questions are appropriate. That is, the evaluation questions are relevant to the evaluation stakeholders and can be answered given

the available data. In addition to important and appropriate evaluation questions, culturally responsive evaluation also involves evaluation stakeholders in decisions about credible evidence. That is, what data will be viewed as valid; necessary for answering the evaluation questions (Frierson et al., 2002; Frierson, et al., 2010).

In the G.A.M.E.S. evaluation, the evaluation questions were developed in collaboration with the program director. To develop the evaluation questions, the evaluators met with the program director to discuss issues she wanted to address in the evaluation. In the meetings, it was evident that the program director was very much interested in issues of diversity, for example, how the program affected different racial/ethnic groups. The program director's interest in diversity was taken into account in the design of the evaluation questions. Evaluators provided the following comments:

In coming up with the questions, we met with the program director to see what kind of information she was interested in getting.

> The evaluation questions were important because we looked at the difference in the outcomes of the different groups, and I think it is important to see how to judge the effectiveness of the program for the different groups and to try to address the difference.

In subsequent meetings with the program director, the evaluators shared the evaluation questions with the program director, who had the opportunity to review the questions and provide feedback to the evaluation team. One evaluator stated: "We went over the questions ahead of time the program director." This feedback was incorporated into the final version of the evaluation questions. Thus, the evaluation questions were carefully considered by the evaluators and project staff (Frierson et al., 2002; Frierson et al., 2010)

In relation to acceptable evidence, the evaluators developed a matrix that depicted the relationship between the evaluation questions, the information required from the evaluation

questions, and planned methods of data collection. Below is a snapshot of the matrix that depicts the linkage between evaluation Question 1, the information required to answer the question, and the methods that were used.

Table 10

Evaluation Question 1 and Methodology

Key Evaluation Question	Information Required	Method	Sample
What is the quality of the content covered during the G.A.M.E.S summer camp, and is this content appropriate for program participants?	Instructional content; content & standards alignment; participant experience of context.	Document analysis and observations; interviews and focus groups of program counselors and participants.	All lesson plans and lab manuals; sample of 13 camp counselors/coordinators and 18 campers.

This matrix provided clear connections between the components of the methodology employed in the evaluation. The matrix was both shared with the program director and was included in the evaluation reports. This matrix not only provided direction for the evaluation, but also ensured that the client was clear as to what the evaluation questions were and the type of information would be gathered to answer the evaluation questions (i.e. observations, interviews, etc.). In regards to the development of this matrix one of the evaluators stated:

> We looked at the evaluation questions with the program director and then from there we laid out how we were going to gather that information. I believe we put that in a chart.

In addition to the matrix above, criteria for judging the program quality was also established. These criteria provided a means for assessing the success of the program and determining which aspects of the program needed further development and improvement. The criteria for judging program quality addressed issues of equity with regards to underrepresented students, in which added to the cultural responsiveness of the evaluation. Of the criteria below,

number four is directly related to underrepresented students. All items of the criteria are included:

1. The program provides high quality instruction that is appropriately connected to the developmental and learning profiles of participants.
2. Program content is of high quality, developmentally appropriate, and congruent with the goal of promoting student achievement in advanced mathematics courses.
3. The program provides meaningful learning experiences for participants.
4. The program increases equity by recruiting children from **underrepresented groups** with access and opportunity.
5. Participants demonstrate meaningful benefits from their experience with the program (e.g., improved mathematics self-efficacy, improved attitude or learning habits, social benefits).

Select an Appropriate Evaluation Design

The evaluation approach should complement the evaluation questions. That is, the methods employed in an evaluation should enable the evaluator to collect the required information for answering the evaluation questions. In culturally responsive evaluation, no one evaluation design is preferred over another: "The evaluation design that one uses does not necessarily need to be elaborate. It just needs to be appropriate for an effective evaluation" (Frierson et al., 2010, p. 85). However, the evaluation design should be appropriate for the evaluation context for what the evaluation seeks to accomplish (p. 85). While no evaluation approach is advocated for in culturally responsive evaluation, these types of evaluations typically employ mixed methods. Mixed method designs generate data by employing both qualitative and quantitative data collection methods.

The G.A.M.E.S. evaluation used a mixed methods design in order to generate a comprehensive, in-depth understanding of the program. Both qualitative and quantitative data collection methods were used to gather data relevant to the evaluation questions. Data collection methods included surveys, interviews, observations, and document review. The program director stated:

> The evaluation activities included surveys to students and parents, focus groups and individual meetings were utilized as evaluation instruments. All of them in my view were equally important to help have a full understanding of the effectiveness of the program.

In this evaluation, the use of a mixed methods design was important because it allowed the evaluation team to gather information on various dimensions of the program. For example, surveys were administrated to gather information on the participants' experiences with the program. In addition, interviews were administered to parents to gather information on the parent's views and perceptions of program. Document reviews provided the evaluators with information on the program's activities and mission, and a mixed methods design was used to develop a deeper and broader understanding of the program (Greene, 2005). The methods employed tapped into different aspects of the program via more than one method. For example, to assess students' attitudes in STEM, both interviews and surveys were administered to the participants. The interview provided first-hand knowledge of the program's impact on the students' experiences and the surveys provided similar information, but also included questions related to other aspects of the program and experiences (i.e. their perceptions about group work during camp).

The use of a mixed methods design was not unique to this evaluation. However, the concept of culture was a critical component. For example, culture was placed at the forefront of the evaluation's design (mixed methods); the goal of the evaluation was to be culturally

responsive in every aspect of the evaluation. For example, information was collected from all the participants, but in some instances, greater efforts were made to collect information from the minority participants to examine if there were any cultural differences in the program participants' experiences. Evaluator comments regarding the design of the evaluation are below:

> I think as far as making it mixed methods, I think most of my evaluations are mixed methods, so I am not sure if there was anything particular in this context or the fact that I wanted to be culturally responsive that said, yeah, we are going to have mixed methods. I typically use quantitative methods first to sort of figure out what issues are the areas of concern and then use qualitative methods to follow up. I think that the [program staff] were very open and they didn't have strong feelings about methodology.

> We did use mixed, and we did gather informative from the participants, and we did try to explicitly get information from the minority participants in terms of their experience in the program. And we did look at the differences of the different groups in terms of their attitudes because we also wanted to measure their attitudes towards STEM.

Select and Adapt Instrumentation

In culturally responsive evaluation, instrumentation is a key component of the evaluation. The instruments employed in the evaluation provide the evaluator with the means to collect data on the program and participants (Frierson et al., 2010, p. 86). In culturally responsive evaluation, evaluators use data collection instruments that are appropriate for the evaluation context. That is, the instruments offer the necessary information for answering the evaluation questions. Additionally, instruments must render both valid and reliable information on the program. In culturally responsive evaluation, the data collection instruments must be appropriate for the target population. That is, the instruments should be culturally responsive to the culture of the program participants. The instruments should also provide both valid and reliable information on the target population.

In the G.A.M.E.S. evaluation, multiple data collection methods were employed to collect data on the program and participants. They included: a) participant surveys, b) interviews, c)

document review, and d) observations. The use of multiple data collection methods provided the evaluation team with the means to generate a comprehensive and in-depth understanding of the program of interest. Moreover, the use of multiple data collection methods provided the evaluators with greater ability to make valid references about the program and participants. Data collected at multiple times in the evaluation also increased the reliability of the data collected in the evaluation. For example, participant information was collected pre- and post program participation via a participant survey. The administration of data collection instruments at multiple times throughout the evaluation, such as the participant survey and observations, increased the reliability of the evaluation and the inferences drawn from the evaluation. In regards to the evaluation instruments, one evaluator stated: "We wanted the instruments to be valid, reliable, cost-worthy, and responsive."

The data collection instruments were not only based on how appropriate they were for the evaluation context, but also on the extent to which they would garner information required for answering the evaluation questions. That is, the instruments would provide information on the program content, pedagogy, and participants' experiences with the program. Below, I describe the relationship between each data collection instrument and the type of information it provided for answering the evaluation questions.

Participant survey. The purpose of the participant survey was to gather information on: a) the experiences of the participants with the program and b) the participant's knowledge of study skills, career planning, and college application procedures. The survey questions were designed to illicit specific information about the quality and appropriateness of the program content and pedagogy. The purpose of the survey was also to access the impact of the program on students' self-assessment of general intelligence, ability in math, and ability in science in

comparison to classmates. The information gathered from the participant survey helped in answering evaluation questions 1 and 2: What is the quality of the pedagogy and content utilized during the G.A.M.E.S. summer camp; is this pedagogy effective; and is this context appropriate for program participants? And In what ways and to what extent does camp participation influence participants' attitudes toward engineering and STEM more generally, as perceived by teachers, parents, and the students themselves?

The survey was given to the participants before and after the program. The survey questions were gathered online from the Online Evaluation Resource Library (CERL's) website (Online Evaluation Resource Library). This library provides professionals with sample evaluation plans, instruments and reports from various NSF projects. The G.A.M.E.S. participant survey was developed using questions from the Douglass Science Institute Program Series: Encouraging Precollege Women to Persist in Math and Science Studies Rutgers University, a National Science Foundation program for women and girls. One evaluator stated:

> I looked at websites, the National Science Foundation has a lot of different outreach programs that were very similar to G.A.M.E.S. and so they have those evaluations and they also have evaluation protocols where they had a list of different instruments that they used with other evaluations in NSF.

The survey covered a range of topics including:

1. General characteristics of student's family background (e.g., those with whom student lives student lives most of the time, occupation of parents or guardian, involvement of these individuals in math or science careers)
2. Student's self-assessment of general intelligence, ability in math, and ability in science in comparison
3. Highest level of education student would like to complete/highest level student expects to complete

4. Student's course history in math and science
5. Student's career plans
6. Student's interest in specific careers
7. Student's attitudes toward scientists and scientific careers
8. Student's participation in class
9. Student's participation in extracurricular activities
10. Information about student's peer interests
11. Student's interactions with others correlated with support for student's interest in math and science

The participant survey, which was developed based on the NSF program above, was specifically designed for the use of programs that serve under-represented populations. Therefore, this survey was an appropriate instrument to use in this specific evaluation context, given the targeted population (i.e. minority girls). One evaluator commented:

> I shared the survey with people in the G.A.M.E.S. program and they did want to change some of the questions, but since I had gotten it from the website, and it had been used before in other NS funded program, I did not change it.

Interviews. The purpose of the interviews was to gather information on the quality and benefits of the program. Parents, program staff, and participants were interviewed in person or by phone. Parent interviews provided information related to the quality of the content and pedagogy provided to the participants, as well as the program's influence on the participants' attitudes and beliefs in science and engineering. Sample questions included:

What are you enjoying about your experiences?

Here at camp, what do you like about your lessons and projects?

Document review. Document review included the collection of information on the program and participants. Documents reviewed included: program manuals, participant data bases, and program materials (i.e. activity sheets, lesson plans, etc.). The purpose of the document review was to gather descriptive information about the program design, recruitment, participant selection and key characteristics of the program participants. Document review also provided information related to the quality of the program content and pedagogy and the type of pedagogical approaches utilized in the camp.

Observations. The purpose of the observations was to capture a descriptive account of the context, program activities, and learners' participation in the camp activities. The evaluators used an observation guide to guide this process. For each observation, the evaluators described the context and setting, documented the key activities, and provided an account of the learner's participation in the activity.

Except for the participant survey, all data collection instruments were designed from scratch to gather information to answer the evaluation questions. For the most part, the data collection instruments remained the same from year to year. However, in the third year of the evaluation, the director requested for additional questions be added to the interview instruments for the participants. These questions sought to elicit the participants' viewpoints about the camp's diversity. Similar questions were added to the parent and program interview protocols.

All the data collection instruments were written in English; however, for the Latina participants, the evaluators served as translators on more than one occasion. That is, the evaluators verbally translated the instruments from English to Spanish during the focus groups sessions. When the participants did not understand a particular interview question, the evaluators translated the questions from English and Spanish. In some cases, the evaluators added

additional context to the questions for the participants to completely understand what was being asked. One evaluator stated:

> The design was culturally responsive in that we did have the instruments translated and we did have bilingual evaluation team members.

Data Collection. In culturally responsive evaluation, both quantitative and qualitative data collection methods are used to collect information on the program. Data collection can occur via questionnaires, surveys, observations, interviews, etc. Culturally responsive evaluation relies on both qualitative and quantitative data collection methods to collect information on the program. However, qualitative data collection methods are preferred because they can provide rich and in-depth information on the cultural context of the program. Additionally, in culturally responsive evaluation, the person collecting the data is also a data collection instrument (Frierson et al., 2010). Data can be poorly collected or analyzed if the person collecting the data does not understand the cultural context of the program or participants.

Based on a mixed methods design, the G.A.M.E.S. evaluation made use of both quantitative and qualitative data collection methods. The evaluation captured the experiences of the participants in the program via the surveys, interviews, and observations. Additionally, the opinions and viewpoints of the parents and counselors were documented via interviews and observations. The use of multiple data collection methods, versus only one, gave the evaluation team a more in-depth understanding of the program under evaluation.

Particularly, the use of observations, a qualitative data collection method, provided the evaluation team with a clear sense of the cultural context of the program. Also, the use of document review, a qualitative data collection method, also provided the evaluation team with substantial information on the participants and program. Participant databases provided demographic information on the participants, including name, race/ethnicity, age, grade level,

school, and hometown. These databases also provided information on the number of participants per camp and participants' status in the program, for example, the number of times they had participated in the program and whether or not they had a received a scholarship (scholarships were offered to students whose participation might have been hampered due to program costs). Document review of program manuals provided information on the program goals, mission, activities, and schedules.

Interviews, another qualitative data collection method, garnered data on the program directly from the participants (i.e. students, parents, and counselors). This method of data collection provided the context for data gathered from the other data collection instruments. For example, the participant survey consisted of both close-ended and open-ended items, and the information gathered from the qualitative methods, the interviews primarily, helped to provide the context for the answers provided by the participants in the close-ended responses.

In addition to qualitative data collection methods, data was also collected in culturally responsive ways. From the start of the evaluation, the evaluators were cognizant of the cultural context of the program. For example, the culture of the participants was taken into account in the design of the instruments. Additionally, the culture of the participants was considered when collecting the data. For example, data on the program participants were collected from the participants; however, the evaluators took steps to ensure that data were collected from the various cultural groups represented with the evaluation. One evaluator stated:

> We gathered information from the participants and we tried to explicitly get information from minorities in terms of their experience in the program.

Information on the program participants was collected via the participant survey and interviews. Participants were randomly and purposely selected to participate in the interviews. Each year, four separate focus groups were conducted with: a) randomly selected participants, b)

repeat participants, 3) African-American participants, and 4) Latina participants. The decision to administrate separate and multiple focus groups was taken in collaboration with the program director. First, the program director was interested in assessing the program's impact on the different cultural groups. Therefore, the administration of focus groups by race would garner specific information relevant to the experiences of the different students. In addition, the decision was taken because both the program director and evaluation felt that it would provide a safe place for the students to voice their opinions, and the environment of a group interview, as opposed to a one-on-one interview, would enable the participants to voice their opinions more freely. In addition, the program director felt that the participants, particularly the Latina students, would be more comfortable if the person facilitating the focus groups was bilingual and of the same racial/ethnic background. The client made the following comment in regards to the use of the Spanish language:

> I know that a lot of Spanish happened between the evaluators and Spanish girls because they did know English. How do I say this, half of them didn't speak very well English, so it was cool to have the evaluators there that could communicate with them and ask them the questions in Spanish. Otherwise, you would have gotten big silences if they did not understand what was going on, or what was being asked. Yes, so I thought that it was pretty cool to have the evaluators speaking the same language that the kids did. That made a huge difference with the Hispanic population.

In addition, the evaluators recognized the importance of the person collecting the data. In qualitative research, the person collecting is also an instrument; if the person is not attuned to the cultural context of the program, the data can be flawed (Frierson, 2010, p. 88). Therefore, when conducting interviews with parents, the evaluators took into account which evaluators would facilitate which interviews and with whom. That is, the evaluation team agreed that it would be beneficial for the evaluators to interview parents from the shared racial/ethnic background to illicit more accurate responses. For example, the Latino evaluation team members facilitated the

interviews and focus groups with the Latina parents and participants. The same was true for the African-American parents and students. One evaluator commented:

> We were culturally responsive in that some of the evaluation instruments were translated, and we had bilingual evaluation team members and African-American evaluators. We also considered who would do what, for example, who would facilitate which focus groups and who would be in the focus groups. We had homogeneous African-American and Latino focus groups, and we talked about that, and so these were all things that I think were design features that we tried to introduce aspects of culture.

In many ways, the evaluation team recognized the importance of a shared lived experience in the evaluation. Therefore, they agreed that the administration of focus groups, in which the evaluator had shared lived experiences, the participants' would be more honest and candid with their responses. In addition, it would help the evaluation team establish better rapport between with the parents if they knew the evaluators were from a shared racial background and/or were able to translate the questions into their native language.

Analyze the data from a cultural lens. In culturally responsive evaluation, it is important to determine the accurate meaning of the data that is gathered in order to make valid judgments about program quality. This requires evaluators to understand the culture of the program and participants. Further, culturally responsive evaluation is sensitive and responsive to cultural factors present within the evaluation. Failure to take into account the cultural context of the program when analyzing data can result in the misinterpretation of the data, and ultimately inaccurate judgments about program quality (Frierson et al., 2002; Frierson, et al., 2010).

In the G.A.M.E.S. evaluation, much care was taken when analyzing the data. Prior to data analysis, meetings between the evaluators and program staff were held to discuss how the data would be analyzed. These meetings occurred early in the evaluation process. Overall, it was agreed that the data, both quantitative and qualitative, would be analyzed via multiple strategies to achieve a complete picture of the program's impact. In addition, the data would be analyzed

from a culturally responsive evaluation approach (Frierson et al., 2010, p. 89), in order to assess the differential impact of the program. One evaluator stated:

> It was analytic theme in the evaluation to look at the data by culture such as race and ethnicity.

A description of the data analysis which took place in the evaluation is provided below.

A key aspect of the evaluation was to assess the change (if any) in the participants' attitudes and beliefs towards STEM. To address this question, a pre-and post survey was administered which included 20 questions that used the Likert scale selections of: 1 = Strongly Agree, 2=Agree, 3 = Not Sure, 4 = Disagree and, 5 = Strongly Disagree. The questions elicited information about the participant's interest in specific subjects and careers, student's attitudes toward scientists and scientific careers, and student's career plans. A list of topics covered in the survey was discussed in the previous section. Sample survey questions from this survey are shown in Figure 2 (below).

Item	Strongly Agree	Agree	Not Sure	Disagree	Strongly disagree
1. I am interested in science.	①	②	③	④	⑤
2. Scientists and engineers work longer hours than most other professionals.	①	②	③	④	⑤
3. Girls who are good in math and/or science are just as popular with boys as others girls are.	①	②	③	④	⑤
4. I can picture myself working in science or engineering one day.	①	②	③	④	⑤
5. I am confident that I can do science/engineering.	①	②	③	④	⑤

Figure 2. Sample questions from the participant survey.

The evaluation team analyzed the quantitative data from the participant survey in subgroups. First, the survey data was analyzed for all participants in the camp. Below is a snapshot drawn from one of the evaluation reports that depicted pretest-posttest comparisons of the survey results for all campers. In the table below, 5 of the 13 of the survey items revealed significant differences. Further, campers were more likely to think that becoming a scientist is a good way to serve humanity after participating in the program (survey item 7).

Question		M	N	SD	P value
I am interested in engineering.	Pretest	1.88	181	.709	.837
	Posttest	1.87		.817	
Students who are good in math and science get picked on, teased, or ignored by other students.	Pretest	4.10	181	.932	.168
	Posttest	4.19		.899	
I am interested in pursuing a career in engineering.	Pretest	2.59	179	.798	.071
	Posttest	2.48		.996	
People who enter math and science careers tend to be "nerds" who lack social skills.	Pretest	4.45	180	.703	.639
	Posttest	4.43		.685	
Scientists and engineers work longer hours than most other professionals.	Pretest	3.09	181	.681	.593
	Posttest	3.12		.814	
Having a career in engineering or science requires too many years of education.	Pretest	3.75	182	.892	.763
	Posttest	3.77		.945	
Becoming a scientist or engineer is a good way to serve humanity.	Pretest	1.90	182	.698	.000*+
	Posttest	1.68		.703	
I have a good understanding of what scientists and engineers actually do.	Pretest	2.28	181	.739	.000*+
	Posttest	1.86		.639	
Scientists and engineers mainly work alone, with equipment instead of people.	Pretest	3.80	181	.770	.005*+
	Posttest	3.99		.925	
Girls who are good in math and/or science are as popular with boys as other girls are.	Pretest	2.10	182	.955	.014*+
	Posttest	1.93		.961	
I am confident I can do science/engineering.	Pretest	1.56	182	.589	.407
	Posttest	1.61		.770	
I can see myself working in engineering someday.	Pretest	2.09	182	.805	.254
	Posttest	2.16		.976	
Working in a science laboratory would not be an interesting way to earn a living.	Pretest	4.14	179	.826	.040*-
	Posttest	3.98		1.025	

Figure 3. Pretest-posttest comparisons of survey results for all campers.

Second, the evaluation team analyzed the survey data by subgroup: race and camp. This method of analyzing the data was purposely used to assess the differential impact of the program on the different groups of participants. The evaluators agreed that one of the methods they used for analyzing the survey data was to disaggregate the data by race and camp in order to assess any statistical differences in the subgroups. Further, data disaggregation was pursued by the evaluation team to garner a more comprehensive outlook on the effectiveness and impact of the program. Evaluators commented:

> We did look to see if there was a difference in the outcomes on the different groups, and I think it is important to see how to judge the effectiveness of the program for the different groups and to try to address that there was a difference.
>
> We looked at the differences of the different groups in terms of their attitudes because we also wanted to measure their attitudes towards STEM and STEM fields. For example, we looked at if they thought they would pursue a career in engineering, so in order to assess different in the different groups, we disaggregated the data by ethnicity.
>
> Below is a snapshot of the average response to one of survey questions by race and camp.

In this figure, all of the participants except for the Caucasian and Latina participants were more likely to agree that people who enter math and science careers tend to be nerds after participating in the camp. Figures like the one below were created for each of the survey questions produced for all campers and the two different subgroups of race and camp. The evaluation reports included the figures in the appendices section.

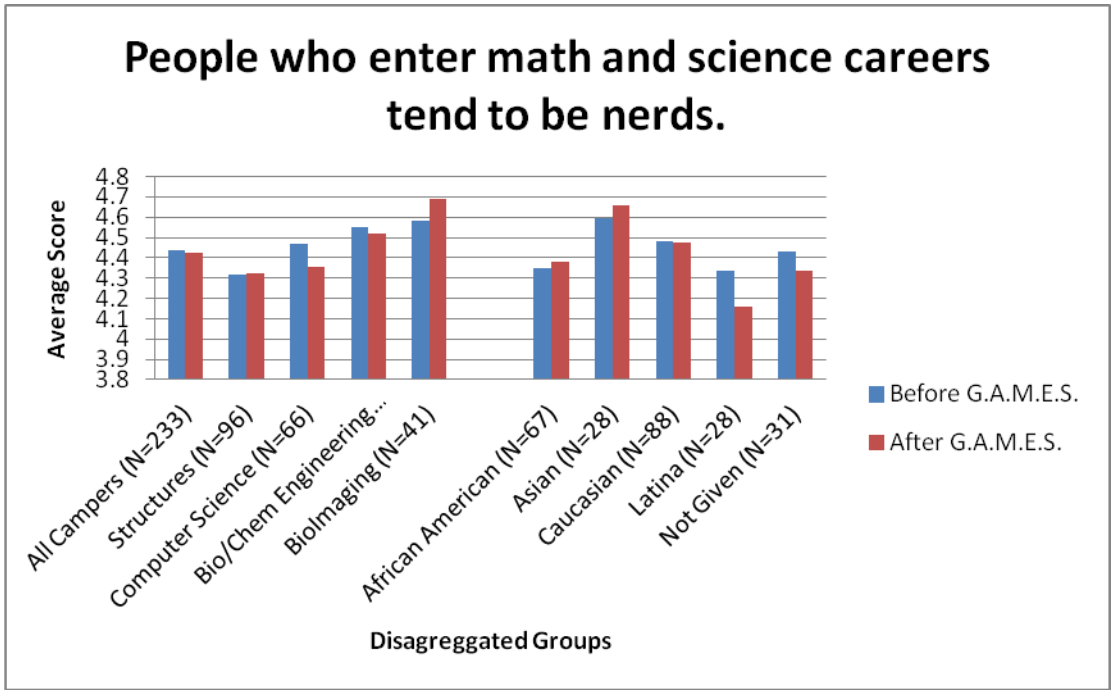


Figure 4. Average response to survey question 4 disaggregated by race and camp.

In addition to analyzing the data by race and camp, the evaluators also considered analyzing the data by socioeconomic status (Frierson et al., 2010). However, this was not done for one reason or another. However, it would have been possible to analyze the data by socioeconomic status given the students who had received scholarships to attend the camp were presumably of lower social economic status than the other campers who did not require financial assistance. The evaluators would have been able to compare the responses of the students who received camp scholarships with those who did not. In comparing the participants' responses, there may have been some statistically significant results which would have given the evaluators greater insight into the experiences of the different groups of participants. For example, the evaluators would have been able to further probe the participants' experiences with and reactions to the program.

The qualitative data gathered from the participant survey was analyzed by employing theme analysis. Themes were developed by subgroups of race and camp. The program director

was very fond of this method of analyzing the data because she was interested in the different responses provided by the different groups of participants to the open-ended questions. In the evaluation reports, the open-ended responses were included for each race and camp. This also gave the program director a better understanding of the difference experiences of the participants in the different camps and by race.

Disseminate and use the results. In culturally responsive evaluation, the evaluation results are disseminated widely and in multiple venues (Frierson et al., 2002; 2010). Evaluation reports and presentation are given to funders, clients, program staff, and participants. The purpose of wide dissemination of the results and in multiple venues is to ensure that the evaluation results are readily available and relevant to program stakeholders. In addition, in culturally responsive evluation, evaluation results are disseminated in ways that will increase the use of the evaluation findings. That is, the evaluation results must be reported in an understandable manner.

In the G.A.M.E.S. evaluation, the evaluation results were disseminated in multiple venues. First, the evaluation results were presented in formal evaluation reports. The reports were first given to the director of the program and then disseminated more widely to program staff, administrators, faculty, and students in the College of Engineering. Because stakeholders were broadly defined in the evaluation, it was important to provide the evaluation to everyone who was considered a potential stakeholder in the evaluation:

We described stakeholders broadly; I mean, part of the stakeholders were program staff, parents, the participants themselves, and other faculty and engineering administrators who kind of sponsored the program.

The evaluation results were also disseminated via oral presentations to the program director, counselors, administrators and participants. The evaluation presentation took place at the completion of the evaluation and was held at a time and place convenient to the program stakeholders. Therefore, the evaluation results were not only disseminated in different formats, but also in multiple venues to increase the use of the evaluation results. The program director felt the presentations were useful to her and the program staff.

> The presentations given to the program staff were pretty good. I mean, they were good because they would hear from the people that were actually doing the evaluations, and they would be able to ask questions, and they would be able to zoom in more in-depth and understand what was going on in the program. For example, I remember the first year we offered the Bioengineering/Chemical Engineering program, that freaking program killed everyone. In the sense that the material was too hard that they kids could not… we lost 50% of the students, and there was a lot of complaints that it was too hard.

Ethical considerations. In culturally responsive evaluation, evaluators are ethical and are bound to two types of ethics when conducting evaluations. That is, procedural ethnics which include those required by Institutional Review Boards, and relational ethnics which include "mutual respect, dignity, and the connectedness between the researcher and the researched and between the researchers and the communities in which they live and work" (Frierson et al., 2010, p. 92). Of the two types of ethics listed above, culturally responsive evaluation is more concerned with the relational ethics because culturally responsive evaluations typical occur in program contexts in which there are underserved and marginalized communities. In evaluation contexts where there are clear power imbalances, evaluators respect and honor the culture of the communities and are "mindful on building relationships of trust and mutual respect" (p. 93).

In the G.A.M.E.S. evaluation, the evaluation team applied for and received IRB approval. IRB procedures were followed by the evaluation team. For example, in terms of data collection, the evaluators provided oral consent to the campers. Further, the evaluators read out loud the

participants' rights. For example, an excerpt of the oral consent for the camper focus group included:

> Your participation in this evaluation is completely voluntary. You are free to withdraw your consent to be interviewed at any time and for any reason without penalty. During the interview, you will be able to interrupt it at any point and for any reason. You are also free to refuse to answer any questions you do not wish to answer. These decisions will have no effect on your future relationship with the school. Your participation in this study involves no anticipated risks beyond those that exist in everyday life.

In addition, the evaluators removed all identifiable information from the data and saved the data in a safe place. In addition to procedural ethnics, the evaluators were also ethical with respect to the relational ethics of culturally responsive evaluation. For example, the evaluators were respectful of the culture of the participants. The evaluators spent a significant time in the program context, learning about and familiarizing themselves with the cultural traditions, habits, and norms represented among the participants (i.e. African American, Latino, etc). In addition, they spent a significant amount time developing trusting relationships with the participants and program staff.

CHAPTER 6

Findings: Strengths and Challenges

The purpose of this case study is to examine how evaluators plan, design and implement culturally responsive evaluation. In the previous chapter, I examined how a culturally responsive evaluation approach was applied to evaluate the G.A.M.E.S. program evaluation. More specifically, I compared the evaluation strategies to the evaluation strategies for conducting culturally responsive evaluation put forth by the 2002 and 2010 User-Friendly Handbooks for Project Evaluation (NSF, 2002; 2010). In this chapter, I discuss some additional findings related to the strengths and challenges the evaluation team experienced in conducting a culturally responsive evaluation.

Strengths

There were a number of strengths of the G.A.M.E.S. evaluation directly related to its use of a culturally responsive evaluation. The strengths of the evaluation were evident throughout: preparation, design, and implementation. In the preparation phase, one strength of the evaluation was the creation of a culturally diverse evaluation team. In the design phase, it was the use of a mixed methods design and multiple data analysis techniques. And throughout the evaluation, a strength was the inclusion of stakeholders in the design and implementation of the evaluation. The strengths of the evaluation are discussed below.

Culturally diverse evaluation team. The first strength is a culturally diverse evaluation team. This responsive team provided the means to better collect, analyze, and interpret the evaluation data (Frierson et al, 2002; Frierson, et al., 2010). In the data collection phase, the evaluation team possessed a set of shared lived experiences which enabled them to relate to and understand the culture of the participants. Further, the evaluation team was familiar with the

cultural norms and behaviors of the participants, which provided them with sensitivity and understanding of the culture of the participants in the evaluation. In the data collection phase, the set of shared lived experiences the evaluators brought with them afforded them with a greater ability to effectively engage with and render accurate responses from the participants. The director of the program was accurately attuned to the cultural diversity of the evaluation and deemed it one as one of the strengths of the evaluation. For example, her comments allude to the critical role the evaluation team played in garnering trustworthy data from the participants.

> The diversity of the evaluation team was very important because it helped the girls feel at ease and comfortable with answering the evaluation questions. They helped them not feel intimidated at all, which played a significant role in the truthfulness of the answers that the girls provided. Middle school is an age where kids are very impressionable and still want to very much please adults. If they do not feel comfortable, they will provide the answers that people will want to hear, and not necessarily the ones that they really believe. By having a diverse team on board, they felt that their needs were taken into account and [they] were willing to be truthful.

> It was beneficial to have a racially diverse evaluation team. I think it would have been hard for the low-income African-American and Latina girls to open up to people they did not perceive as equal to them.

In addition, the evaluation team's shared lived experiences also provided them with a greater ability to analyze the data more accurately. In particular, data garnered from observations revealed cultural differences in participant engagement which was largely attributable to differences in communication styles (Frierson, et al., 2002; Frierson, et al., 2010). Because the evaluation team was familiar with and had an understanding of the communication styles and culture of the participants, they were better able to identify and explain the participants' behaviors in culturally specific ways. Two prime examples that demonstrate the evaluation team's ability to identify and interpret cultural nuances that impacted the evaluation data from the observations are described below.

In the data collection, the evaluation team observed that the Latina participants' level of engagement varied and was dependent on the type of camp activity. For example, the Latina participants were less engaged in the activities that required them to communicate verbally in English. For example, lectures and demonstrations required participants to provide feedback on the material through group work. In group work, Latina participants were observed as less engaged in comparison to the other participants. Further, Latina participants rarely asked questions or commented on the material. However, in the hands-on activities, the Latina campers exhibited different participation levels. For example, in the Structures Camp, which requires campers to build a wooden boat by applying engineering concepts, the Latina participants were observed as being actively engaged in the activity, in some cases more than the other participants. The Latina participants built more effective boats and often completed the tasks at hand faster. In this example, the evaluators were familiar with the culture of the participants. Further, they recognized the cultural barrier (language) was a negative factor. Therefore, the evaluators were able to identify and interpret such cultural nuances effectively.

In addition to the communication styles observed in the Latina participants, the evaluators also recognized certain cultural nuances in the African-American participants. Similar to the Latina participants, the African-American participants were also observed in different ways during the tasks at hand. For example, in the lectures and classroom demonstrations, the African-American participants were generally less engaged in comparison to other participants. For example, many of the African Americans participants sat in the back of the classroom without very little attention to the tasks at hand. Similar to the example above, the evaluators were able to discern certain cultural barriers that impeded the engagement of the participants. For example, the African-American evaluator was familiar with the types of behaviors exhibited in

African-American culture that may have been interpreted by some as negative or disrespectful. The African-American evaluator provided the following comment in regards to the behaviors of the participants:

> I felt like I had an understanding with the students, well at least most of the students that I taught this coming year, and most of the students that I taught were like the African girls in the program. So I felt like I had a good understanding of them like when I watched their behavior, what seemed to be very similar to the students that I taught previously.

Another evaluator who was not African American provided the following comment in reference to the cultural insights of the other two evaluators:

> I remember when she [the African-American evaluator] would say that the African-American girls sat in the back of the room and were loud and that connected to some of her own experience, right, and I remember when [the Latina evaluator] said that the Latina girls were good with doing things with their hands [i.e. building the boats] and that is an insight that I [as a Caucasian] evaluator would not have had, and that is something the evaluation team, because it was ethnically diverse, brought to the evaluation.

The examples above demonstrate the ways in which a culturally diverse evaluation team strengthened the validity of the results. The evaluators had a collective understanding of the culture of the participants: cultural norms, values, language, and behaviors (Askew et al., 2012). Because the evaluation team possessed a shared lived experience, they had a greater ability to not only identify and but interpret cultural differences in the data. Further, the evaluation team achieved multicultural validity in that they accurately captured and interpreted the cultural experiences of the participants (Kirthart, 1995).

However, the evaluation team recognized the importance of cultural competency. Cultural competence refers to the "the education, abilities, skills and experience appropriate to undertake the tasks proposed in the evaluation" (AEA, 2004). While the evaluation team was culturally diverse, they acknowledged that there may have been some limitations in their knowledge. For example, while their experiences were similar to the participants, some differences may have been present between the evaluators and participants that hampered their

ability to fully understand the culture of the participants. Because race/ethnicity is only one facet of culture, other cultural factors such as socio-economic status must be accounted for when attempting to the fully understand the culture of a specific group of people. Evaluators provided the following comments in regards to their shared lived experiences:

> You cannot always assume that because I am Italian I know how all Italians feel. You cannot say, "well because she is Latina, you know how Latinas are," or "I know how all Latinas behave like because she is Latina." You have to understand the person as a person, not only as a Latina.

> You are never quite sure if you are interpreting things quite correctly. You do get direct information from the participants. You have to ask the right questions. I am African American, but I did not grow up in an inner city neighborhood or in that type of situation at all. I probably have more in common with the suburban life in some respect. In some ways, I was able to interpret the inner city black and Hispanic group experience, but I do not know how accurate I was at that.

In addition, the evaluators also spent a significant amount of time collecting background information in the context of the evaluation to expand upon the evaluation team's collective knowledge on the culture of the program. The evaluators acknowledged that it was important for them to develop a greater understanding of the culture of the participants beyond their shared lived experiences. As such, while culturally responsiveness was a goal of the evaluation, the evaluators also acknowledged that they also needed to gain a sense of cultural competency. As Frazier-Anderson et al. (2011) noted, "A certain level of cultural competence is necessary to successfully conduct evaluations within majority African American settings" (347). Therefore, to complement their shared lived experiences, the evaluators also gained first-hand knowledge of the culture of the participants in the program through observations and other data collection methods. One evaluator commented:

> One of the most significant aspects of culturally responsive evaluation is really to be there to understand what is going on, but also to understand the culture of the program, participants, organization, or institution. And I think that both of those things imply time on site and lots of interaction with the participants and participants.

> I think you have to [have] cultural competency, or at least some level of cultural competency, before you can conduct a culturally responsive evaluation because you have to know the elements of culture that make a difference, and you need to attend to these. But they are iterative. I think that you have to have some level of competency to do culturally responsive, and I think [if] you do culturally responsive it helps you with your cultural competency.

In addition to shared lived experiences and cultural competency, the evaluation team also possessed other skills that were valuable to them throughout the evaluation process. The evaluation team was highly trained and had experience implementing culturally responsive evaluations and other types of evaluation approaches that also account for culture in the evaluation, such as EVEN evaluations (Greene, et al., 2006). In addition, the evaluation team also had a certain level of expertise in conducting evaluation in diverse evaluation contexts both internationally and domestically. One evaluator had evaluation experience in Africa and another in Chicago Public Schools, which are both culturally diverse and highly complex evaluation contexts politically, historically, and economically. Evaluators commented on their evaluation experience:

> I did some evaluations in Africa where it was a totally different culture, how people communicate, the time it takes to do things, what's valued, what things are important, the ideas of goals and positivism, it's a very foreign culture, and I think in order to be an effective evaluator, you have to suspend all of your ideas and give time and energy to what other people value and believe.
>
> Working in Chicago, you know Chicago Public Schools is a culture, there are different people with different cultures in Chicago Public Schools, and Chicago public schools is a culture within itself. A lot of people crash and burn because they do not get the culture.

The collection of shared lived experiences, skills, and experiences the evaluators brought with them helped to ensure the evaluators were competent and responsive to the cultural context of the evaluation. That is, the evaluator possessed a fundamental understanding of the culture of the participants, including the cultural norms, experiences, traditions, and behaviors. As such, the evaluation was attentive to the participants' cultural values in the form of cultural nuances. The

attention paid to cultural nuances in the analysis and interpretation of the evaluation results provided the evaluators with a more accurate analysis of the data from a culturally responsive evaluation approach. Therefore, in more than one way, the diversity of the evaluation team provided greater multicultural validity (Kirthart, 2005) to the evaluation results. In addition, the data was collected more effectively because the evaluators accurately interpreted the impact of culture in the evaluation; in culture impact program outcomes this aspect is key to achieving multicultural validity (Frierson et al., 2002; Frierson, et al., 2010). The client provided the following comment in relation to the evaluation:

> This [evaluation] was a well-crafted design instrument that speaks volume to the professionalism and degree of understanding of the underlying issues that [the evaluators] had of the goals of the program**.**

Mixed methods. The use of a mixed methods design in the evaluation was a strength of the evaluation. While no one method is preferred over another, culturally responsive evaluation advocates for the use of mixed methods (Frierson et al., 2002; 2010). Mixed methods studies with complementarity seek a deeper and broader understanding by using methods that tap into different aspects of the same phenomena. Further, researchers mix methods for the purpose of triangulation when different methods are used to examine the same phenomena for the intention of convergence (Greene, 2007). In the G.A.M.E.S. evaluation, mixed methods were used for the purpose of complementarity and some data triangulation.

First, the use of mixed methods for the purpose of complementarity provided evaluators with a deeper understanding of the program. Because mixed method designs call for the use multiple data collection methods, the evaluation team utilized a mix of quantitative and qualitative methods. The collection of both qualitative and quantitative data provided the evaluators with better understanding of the program and participants. Further, the use of multiple

methods provided various means to assess the same phenomena. For example, the phenomenon of student attitudes in STEM was assessed through various methods: participant surveys and interviews. The quantitative data rendered from the participant survey reveled critical data on the students' attitudes and behaviors toward math and science. Similarly, the qualitative data from the interviews rendered pertinent information also relevant to the participants' attitudes toward STEM, as perceived by the students. The combination of data, qualitative and quantitative, provided the evaluators with a broad and well-balanced understanding of the program and in some instances, data triangulation was achieved.

In addition, the collection of data from multiple methods beyond the participant survey also rendered data relevant to the purpose of the evaluation: to assess the quality of the evaluation. The use of multiple data collection methods provided the evaluators with greater ability to assess the program. The use of document review and interviews enabled the evaluation team to assess student outcomes via multiple methods. For example, the use of document review in the evaluation rendered critical information relevant to quality of the content, pedagogy, and experiences of the participants (Evaluation Question 1). In addition, the use of interviews rendered critical information relevant to the influence of camp participation on the students' attitudes toward engineering and STEM more generally (Evaluation Question 2). The combination of both types of data provided the evaluation team with a robust understanding of the impact of the program on the different participants. In addition, through the inclusion of the culture in the evaluation, the evaluators were able to identify differences in program experience by the different groups in the evaluation. When asked, "What are the kinds of evaluation activities that occurred? Of those which do you see as the most important? Why were these important?" the program director provided the following comments:

> Surveys to students and parents, focus groups and individuals meetings were utilized as evaluation instruments. All of them in my view were equally important to help have a full in-depth understanding of the effectiveness of the program
>
> We got the quantitative data, and that was quite useful. Then the evaluators had been doing the interviews, and that was useful as well because the evaluation could go into the details as to what the girls said.

In conclusion, the use of a mixed method increased the validity of the evaluation results. First, the evaluation design rendered a deeper understanding of the program. Second, it provided the evaluators with the greater ability to assess program outcomes through multiple means. In addition, the evaluation tapped into different aspects of the same phenomena, and in some cases achieved data triangulation.

Instrumentation and analysis: Another strength in the evaluation was the selection of instrumentation and the means of analysis. First, the instrumentation was valid and reliable and rendered accurate data. In the G.A.M.E.S. evaluation, the evaluation team selected instruments with previous history. For example, the evaluators adapted an existent survey for use in the evaluation. The existent survey included a survey given to assess student attitudes in STEM in various NSF projects. The instrument was not only reliable and valid, but was normed for the various populations served in the G.A.M.E.S. summer camp: underrepresented minority students.

In addition to the use of existent instruments, other instruments used in the evaluation were carefully crafted to "capture the kind and type of information needed to answer the evaluation questions" (Frierson et al., 2010, p. 86). For example, the interview protocols were developed with the intent to gather information that would help render data to answer the evaluation questions. The evaluation's team attempt succeeded; each evaluation instrument provided an additional layer of information. In addition, the instruments were developed to be

culturally sensitive. In some cases, the questions in the interview protocols were translated beforehand, or the evaluators translated the interview questions from English to Spanish.

In addition to instrumentation, data analysis was also one of the strengths of the evaluation. All data rendered in the evaluation was analyzed with "considerable sensitivity to, and understanding of, the cultural context of the program" (Frierson et al., 2010, p. 89). In addition, data was analyzed quantitatively (i.e. descriptive and inferential statistics) and qualitatively (theme and content analysis); however, the emphasis was on analyzing the data through a cultural lens. Quantitative data was analyzed by race and camp; attempts were made to analyze the data by socio-economic status. In addition, stakeholder input was considered on how to analyze the data; their requests were met (Manswell-Butty, et al., 2004).

Data analysis of the quantitative data by subgroup revealed significant differences between the participants. For example, significant differences in the students' attitudes towards STEM revealed that the participants experienced the camp in different ways. Again, the purpose of data disaggregation was an attempt to look at the differential impact of the program on the student outcomes, and some of these differences were revealed in the data analysis. For example, the first year of the evaluation, the data suggested that African-American participants entered the program with a much stronger idea of what a scientist was and much stronger feelings that they wanted to be a scientist. Further, this insight was particularly useful when examining differences in the students' outcomes and judging the quality of the program. Evaluators provided the following comments in relation to data analysis:

> I felt like we did a pretty good job of explaining the experience of the minority students. I really enjoyed the quantitative piece because we were able to find many significant differences in the data. A lot of times when you disaggregate the data by race, people do not necessary try to do anything to close the gap. I felt like with this particular evaluation, they did try. I had a good time doing it, and I thought we did a good job with it, and it was actually used.

> I think some of the strengths was the fact that we did pre-test and post-test, and we were able to identify significant differences and subgroup differences that the program worked differently for the white kids than it did for the African American and Latino kids. I think that was a major finding, which made the developers think differently about the program.

Stakeholder engagement: The engagement of stakeholders in the evaluation also strengthened it. The program director was directly involved in the evaluation, in the development of the evaluation questions, and in the dissemination of the evaluation results. Her participation in the evaluation was particularly useful in its design. The program director provided sound advice to various components of the evaluation, including the purpose, design, methods, and instruments. In addition, the program director reviewed and approved every aspect of the evaluation.

In addition to the evaluation design, the participation of the program director in the evaluation approach also strengthened its credibility. For example, the evaluation was more credible to the stakeholders because they were directly involved in the evaluation process: "Indeed proponents of collaborative evaluation argue that one of its great benefits is increased likelihood of crafting a credible evaluation, with results and recommendations that stakeholders, by virtue of their participation in the process, not only understand but are also prepared to address" (Askew, et al., 2012). In the evaluation, program stakeholders were involved in every phase and stage of the evaluation. For example, evaluation methods and instruments were developed in consultation with the program director of the evaluation. In addition, the program director approved the evaluation plan and participated in the use of the results. Therefore, through the program director's direct input in the evaluation, it was more credible to her and to other program stakeholders.

The program director's participation in the evaluation also helped to promote better use of the evaluation results "A second assumption fundamental to these approaches is that

stakeholder engagement and program staff participation in the evaluation process is a necessary prerequisite for evaluation *utilization*, also viewed essential for program success." (Askew, et al., 2012).The program director also agreed that the evaluation provided her with useful information. She described the strengths of the evaluation in the following way:

> The evaluation strengths were: a) it provided us with data about students' attitudes about the subject matter; b) it provided us with data about the students' self-efficacy and self-concepts beliefs; and c) it provided us with data about students' understanding of the materials being covered in the program and their perceptions of their difficulties, which help us recalibrate our classes.

The evaluators also agreed that the evaluation also provided useful information to the program stakeholders. The evaluators agreed that one of the strengths of the evaluation was the program directors' direct engagement in all of the evaluation activities, which increased the use of the evaluation results. An additional strength of the evaluation was that the program director was not only supportive of the evaluation, but also very open to the evaluation and its use of a culturally responsive evaluation approach.

The evaluation feedback, both formative and summative, was used to improve the program. First, formative feedback focused on strengths and weaknesses of the program. Summative feedback focused on the program's impact. The first evaluation found that there was very little diversity among the camp counselors and participants; one of the recommendations provided in the evaluation was to increase camp diversity and provide diversity training to camp counselors who had very little experience with diverse students. One evaluator provided the following comment:

> We tried to look at the data from the perspective of the camp counselors because most of the counselors, or all of the counselors, were women but there was not much diversity. And many of them had very little experience, I would say, with culture, and so I do not think that they had a lot of insight and reflection into it. They were sort of at the color blind stage where they were like, "I treat all children the same." That evolved over time; the second year the [program director] tried to pick more diverse counselors and students.

> It was often in terms of internationals. I am sure she tried to attract everyone that she could, but she did bring in more Indian students and other people from other cultural perspectives and so that maybe was an artifact of the evaluation.

Another noteworthy finding was in the second year of the evaluation. Based on the recommendations of the first year evaluation, the program recruited a larger number of racial/ethnic minority students. The data revealed that the African-American girls came in with a weaker idea of what a scientist was or that they wanted to be a scientist. As a result, the African-American students experienced significantly higher gains in the pre- and post-test attitude survey toward math and science. Therefore, the evaluation data reported significant differences based on the different backgrounds of the participants and gave useful insight into program improvements. Based on the evaluation results and recommendations, the program director made improvements to the program. In subsequent years, the program director provided background information on some of the activities to the participants. One evaluator commented:

> What they tried to do was to provide some background information before the kids came to camp, and they also tried to have a day before camp started to give them some background activities.

The evaluation was also useful to the program stakeholders in other ways. The evaluation was used in funding applications for the program funding. Further, the evaluation was also used to promote the program and others within the College of Engineering. Most importantly, the overall evaluation was useful because it provided program stakeholders with information on how to improve the staff. Therefore, the evaluation assessed the quality of the program content, pedagogy, and experiences of the participants. Additionally, it provided stakeholders with data relevant to specific underrepresented groups in the evaluation and provided specific recommendations for each group. Additionally, it also provided professional space to discuss sensitive issues, which would have otherwise been overlooked with the diversity issues:

> Several of the counselors were relieved that some of the differences they observed in the participants' backgrounds was addressed in the data. I think it was a positive way of saying that the kids are coming with really different backgrounds. Where the counselors, that might have been seen as racist or negative had they said the African American kids do not come in with a very good idea of what camp is about, I think it made, the data revealed it in a way, the way that I think made it safer to talk about it, and we tried to do that.

Challenges

In addition to successes, the evaluation team also experienced some challenges common to culturally responsive evaluation. The challenges experienced in this evaluation were related to resources and not to the complex nature of the program in which culturally responsive evaluation take place. The challenges experienced in this evaluation were related to resources needed to conduct a successful culturally responsive evaluation. According to Frazier-Anderson et al. (2012), there are a number of evaluation resources needed to effectively complete a culturally responsive evaluation, and they include:

1. Funding (the money required to pay for the evaluation)
2. Technology (the type of technological equipment needed to fulfill the contractual obligations for the evaluation, such as computers, printers, video-and audio-recording devices, and transcription for machines and services)
3. Materials (the tools and equipment needed to complete the evaluation, including the software needed for survey tool development and data analysis)
4. Physical environment (the headquarters where evaluators are to be stationed during the course of the evaluation) (p. 367)

Further, challenges common to culturally to responsive evaluation were not an issue in this evaluation, and they included: a lack of representation among key stakeholder groups (Greene, 2000; Frierson et al., 2002; Thomas 2004; Ryan et al., 2007; Frierson, et al., 2010), a shortage of

staff to fulfill all evaluator roles (Manswell-Butty, 2004), competing agendas between schools, districts, and communities (Thomas, 2004), and the complex and dynamic nature of urban school settings (LaPoint & Jackson, 2004). Below I discuss the challenges experienced in this culturally responsive evaluation.

Labor intensive. The evaluation was very labor intensive (Thomas, 2004; Manswell-Butty, et al., 2004). Before the evaluation, evaluators spent a significant amount of time preparing for the evaluation. Thomas (2004) noted that at least one year "is needed to conduct start-up activities, such as learning about a school's physical structure and functioning, the school's personnel and stakeholders, and the school's culture and climate and to collectively plan interventions and collect baseline data" (Carroll, LaPoint, and Tyler, 2001). The evaluators collected data on the program via document review and interviews. They also held meetings with the program director to plan and design the evaluation. Based on the meetings with the program director, the evaluators designed a robust evaluation plan that included a description of the evaluation's purpose, context, design, and data collection methods. In addition, the evaluators developed a series of evaluation instruments from scratch (i.e. group and interview protocols). During the evaluation, the evaluation was even more labor intensive. Evaluators administered a series of observations of program activities and campers. In addition, the evaluators administered a number of informal and formal interviews with program staff and participants. After the evaluation, evaluators analyzed the data and wrote an extensive evaluation report. In addition, evaluators prepared and administered oral presentations to program staff and administrators on the evaluation findings, which also required a large amount of time and effort on the part of the evaluators to design a formal presentation of the results and recommendations. One evaluator stated:

> Just collecting data from people I think was my biggest challenge. And you just have to be persistent about it and try to make sure you them a place to do I and collecting right then there. It was a big challenge for this evaluation, especially because we collected so much data all at once.

Limited resources: In addition to the amount of labor required to design the evaluation, the evaluation also rendered very limited resources. The evaluation budget was very small; only enough to hire graduate hourly assistants. Grad hourly students had other obligations to attend to, which not only limited the amount of time they were able to dedicate to the evaluation, but also the evaluation efforts. In addition, the amount of funding set aside for the evaluation also limited the amount of hours the evaluators spent on the project. This meant that the evaluators were only able to spend a prescribed number of hours collecting and analyzing the data, as well as writing the evaluation. Increased funding for the evaluation would have been beneficial because it would have provided the evaluators with more time to collect and analyze the data, as well as present the evaluation to other stakeholders of the evaluation, for example, parents and members of the community. Evaluator comments on the evaluation budget are included below:

> The evaluation was not well funded. We only had enough money for graduate hourly. We cut back on the budget in terms of the hours from the first year. We definitely had more coverage over the camp the first year than we did in the second and third years. And it would have been nice to actually have interviews with students to follow up on what was said on the survey, to see how those attitudes were maintained later. That would have been nice, particularly with minorities when they get back home.

Time intensive: Time was also a challenge for the evaluation. Time was an issue particularly for data collection and analysis. The program lasted one week, which limited the evaluators' efforts to collect data. Additional data collection on the history of the program would have strengthened the evaluation team's understanding of the cultural, social, and political dynamics of the program (Butty, Ried, & LaPoint, 2004). The time span also limited the amount of time spent on-site on data collection. The evaluation team administered a number of

instruments throughout the program; however, more time would have provided the evaluation team with greater attention to detail. Even though substantive efforts were made to collect data as efficiently as possible, there were some program activities that interfered with data collection. In addition, after the evaluation was complete, the evaluation team spent countless hours analyzing data from over 250 participants, as well as analyzing data from one-on-one interviews and focus groups. After data analysis was complete, the evaluation team put together an evaluation report which was substantive in nature and included a description on the evaluation, and more importantly, a list of the evaluation findings and recommendations.

Limitations

Below I describe general limitations to evaluation. These limitations are not specific to culturally responsive evaluation; however, they would have strengthened the evaluation.

Data quality. The first limitation is related to data quality. For example, the evaluators did not collect longitudinal data on the participants. Longitudinal data on the participants' attitudes and course taking patterns post-participation would have increased the quality of the evaluation design; however, this design feature was not part of the evaluation. The program director agreed:

> The evaluation was useful for me when I was there. After a while, I would have liked some longitudinal data to see how the trends played out. I know there were differences in the time the evaluation took; however, I would have liked to have known if the program made a difference longitudinally, for example, year to year. Like every year, did the students respond the same or differently, and why.

In addition, the evaluation could have been strengthened via follow-up with the students. The first year of the evaluation, a follow-up survey was mailed to the participants and the evaluation team received a low response rate. Therefore, this type of data collection was not

pursued in the subsequent years of the evaluation. One evaluator stated that more follow-up with the students would have been beneficial for the evaluation:

> I think one thing I would have like to have done was do more follow-up with the students because some of the kids were going back to like rich suburban schools where all this would be very reinforced, and some for girls were going to back to not very good schools. And so I think sort of being able to track that would have been a good thing.

Also, more follow-up would have rendered more data on the students' experiences. While the surveys and focus group provided information on the student's experiences at the time of the program, more follow-up with the students post-participation would have given the evaluators a better understanding of the program impact, before and after the program was completed by the students. One evaluator stated:

> Tracking students would have given us information on the different experiences of the kids, and so it would have been nice to see how it played out once the kids got back to their culture, and we didn't really have the capacity to do that.

Additionally, the client agreed that more follow-up with the students and longitudinal data would have increased the quality of the evaluation:

> I would have like to have tracked the students, perhaps until after they graduated from high school, and then ask them 'How did G.A.M.E.S. influence your career selection?'

In addition, an experimental design with a control would have also strengthened the design of the evaluation. The goal of the evaluation was not to have an experimental design; however, it would have been possible to compare the students who completed the program to students who applied but did not get accepted for one reason or another (i.e. maximum capacity). The program director commented:

> The other thing is, it would have been nice to have had a control group to compare the kids who got in to those who did not. We could have compared the attitudes of the kids who did not participate in the program to other students.

Program leadership. After the evaluation was completed under the leadership of the program who requested the evaluation, the evaluation was terminated after a new program director was appointed to the position. The original program director left at the end of the third evaluation, as she was offered a new position at a different institution. After the original program director, a new director was appointed to the position. The new program director, for one reason or another, did not show interest in the evaluation. As a result, the evaluation was not continued in subsequent years under the new program director's leadership. One evaluator shared the following comment on the change of leadership and its impact on the program:

> I think the most negative thing about the evaluation was that when the leader [program director] left and when the leadership changed. Even though we had really good data to show that this was a really good program, it was not enough to sustain the program, and I do not think that it has an emphasis on diversity at all in how the program continued.

In addition to not continuing the evaluation, the new leadership also did not take into account the evaluation of the previous years. One evaluator stated:

> I think one of the outcomes that I feel sad about is that the evaluation was not taken into consideration in how the program continued. In fact, after the original program director was gone, there was a backlash to kind of erase her part of the program and do it another way, and I think that happens in organizations.

While the data and evaluation in general proved to be very useful to the original program director, this data was neither considered nor taken into account by the subsequent leader. While there were parts of the evaluation that would have also proved useful to the new program director, the evaluation's recommendations were not implemented or enforced. This was a limitation of the evaluation and also for the new leadership in implementing the program in future years. One evaluator stated:

> It was sad that they did not look at the data and then say what facets of the program should we keep or what changes should we make?

In addition, the last year the evaluation was completed, the evaluation team provided the new director with the evaluation report from the previous summer's evaluation. However, the program director did not show interest, for one reason or another, in discussing the evaluation results with the evaluation team. Every year, evaluators presented the evaluation results to program stakeholders in a formal oral presentation, in which all members of the Engineering community were invited to attend, including faculty, administrators, and students. While the evaluation team sent numerous emails to schedule this presentation, the new director did not respond to the invitation. Further, while the evaluation report was provided, the evaluation results were not discussed or further explored through a formal meeting. When asked about challenges in the evaluation, one evaluator provided the following comment:

> I think at the very end, I remember the original director left in August and we had the report ready in like September or October, and I remember we did not know who to give it to, or report to. I mean I do not think we ever closed the book; you remember we tried to get a meeting, and we finally gave them the report, but we did not meet.

In addition, the client also noted that the evaluation was no longer actively purposed after she left. In addition, she felt that while the new leadership might be interested in the evaluation, they had not sought nor have secured funding for evaluation or paid evaluators to evaluate the program. The program director commented:

> I think they see the benefits of the evaluation, where I see the big difference is that nobody, in my case, I took G.A.M.E.S because they told me to. So I took over years and what I felt so much love for the kids that would do whatever they wanted, so I was willing to write the grants, to get the money. I think the issue is nobody wants to write the grants. If they don't write the grant, they don't have the money to pay you guys [evaluators].

Representation. An additional challenge of the evaluation was a lack of representation among key stakeholders groups in the evaluation. While the evaluators attempted to involve all key participants in the evaluation – parents, program staff, participants, faculty, and

administrators – parents were less involved in the evaluation activities. Parent interviews were the only method in the evaluation that assessed parents' views and opinions about the program, and level of satisfaction. While the interviews were helpful, and provided relevant and important data for the evaluation, gaining an understanding as to why the parents enrolled their child in the program would have also been beneficial for the evaluation. In addition, collecting data from diverse parents on various topics, not only why they enrolled their child in the program, would have also been meaningful for assessing the differential impact on the program on the different students. Evaluators commented:

> Well, the thing that I always kind of wanted to do was talk to the parents because I felt l like it would have been good to really understanding of motivation for putting their kids in the program and their criterion for selecting a program.
>
> We did try to engage them when they were on site, but they were on site for just a small amount of time. I don't know that we got as many stakeholder perspectives in terms of parents as we could have, and particularly from diverse parents.

While the G.A.M.E.S. evaluation reflected the key dimensions of culturally responsive evaluation well, there are other ways the evaluation could have been more culturally responsive. For example, the evaluation team could have selected an evaluator to serve on the team from an Asian American cultural background, or the evaluation team could have designed an evaluation that would have captured the experiences of the parents more completely. However, despite some limitations in the evaluation, there are many ways in which this evaluation serves as an exemplar of the use of culturally responsive evaluation. Therefore, given the constraints the evaluation was situated in, common in culturally responsive valuation, for example, limited time, resources, etc, the evaluators succeeded at implementing a culturally responsive evaluation.

CHAPTER 7

Findings: Discussion and Recommendations

The purpose of this case study was to examine how evaluators plan, design and implement culturally responsive evaluation. In the previous chapter, I analyzed the results of the study. In this chapter, I summarize and discuss the main findings. In addition, I discuss recommendations for future research and conclude with some final thoughts.

Summary of Main Findings

Data gathered for this study yielded several main findings. First, there was evidence of the use of a culturally responsive evaluation approach. Further, the evaluation employed the culturally responsive evaluation strategies outlined in the *User-Friendly Handbooks for Project Evaluation* (NSF, 2002; 2010). Further, the strategies were employed throughout the evaluation process and in the three main phases of the evaluation: preparation, design and implementation. Through the use of culturally responsive evaluation strategies, the evaluators successfully addressed the cultural context of the program within the evaluation. In addition, the evaluation employed culturally responsive evaluator strategies that directly aligned with and were responsive to the culture of the participants (i.e., culturally sensitive evaluation instruments). Thus, the data gathered from observations, document review, and interviews provided evidence of the use of culturally responsive evaluation.

Second, there was evidence of challenges in the implementation of a culturally responsive evaluation. The first challenge was that the evaluation was labor intensive. Evaluators spent a significant amount of time in the evaluation context. The most labor intensive part of the evaluation was the collection and analysis of data. Evaluators employed a number of data collection methods: surveys, interviews, document review, and observations to gain a full

understanding of the program. In addition, the data entry and analysis of 250 surveys required a significant amount of work. Another notable challenge was time. For example, the evaluation took place over a one-week period, and while evaluation reporting fell into the fall semester, there was only a limited amount of time to gather data in order to fully understand the cultural context of the program before the evaluation. Further analysis of the history, goals, and structure of the program would have been beneficial to the evaluators' understanding of the culture of the program (Manswell-Butty, et al., 2004). In addition, additional time would have allowed the evaluation to perform a deeper analysis of both the quantitative and qualitative data.

While the data yielded challenges relevant to culturally responsive evaluation, there were notable strengths in the use of a culturally responsive evaluation approach. For example, the use of culturally responsive evaluation increased the validity of the evaluation results. The evaluation team was able to accurately analyze and interpret the evaluation. In addition, the evaluation team's shared lived experiences provided them with a greater ability to accurately depict the cultural experiences of the participants (Kirkhart, 2005). In addition, the evaluation design was of high quality, which included multiple data collection methods and culturally responsive evaluation instruments. The instruments not only yielded information that helped to answer the evaluation questions, a key feature of culturally responsive evaluation, but also produced valid results.

The use of a culturally responsive evaluation approach also yielded positive benefits for the credibility and utility of the results. For example, the evaluation was more credible to the stakeholders because they were directly involved in the evaluation process, which is a key component of culturally responsive evaluation (Frierson et al, 2002; 2010). The evaluation plan, methods, and instruments were developed in consultation with program stakeholders. In addition,

the program stakeholders had input on the data analysis and dissemination. Increased credibility on the part of the program stakeholders led to greater use of the evaluation results (Askew et al., 2011). The evaluation was useful in multiple ways: First, it was used to secure future program funding. In addition, it was used to promote programs within the College of Engineering. Lastly, the evaluation was useful because it provided program stakeholders with evaluative feedback relevant to their evaluation needs.

There are some notable points of discussion relevant to the main findings of the study. For example, the success of the G.A.M.E.S. evaluation cannot be entirely attributed to efforts of the evaluation team. For example, the circumstances which surrounded the evaluation context provided a supportive context for the use of culturally responsive evaluation. In the G.A.M.E.S. evaluation, the program director's openness and support towards the use of a culturally responsive evaluation assisted in the proper facilitation of a successful culturally responsive evaluation. Moreover, the program director's request for the evaluation that stemmed from own her interest in cultural diversity also increased the likelihood of the effective use of a culturally responsive evaluation approach. Had the program director not been open to the use of culturally responsive evaluation, or had the program director's interests not been aligned with this type of evaluation, perhaps the evaluation would have not been as successful as it was. However, the evaluation team's efforts to conduct a culturally responsive evaluation should not be undermined, as they helped produce a high quality culturally responsive evaluation. However, in many ways, the G.A.M.E.S. was even more successful because the evaluation context was supportive of a culturally responsive evaluation.

In addition, the evaluation's success can also be attributed to the program context in which the G.A.M.E.S. evaluation took place in. In many ways, the cultural context of the

program lent itself to the use of a culturally responsive evaluation approach. The program context was highly diverse, which involved a large number of participants representing various cultural backgrounds, including African American, Latino, and Asian Americans. Additionally, socio-economic diversity was represented among the participants. Evaluations of programs that involve ethnically diverse participants and stakeholders present the need for the "creation of multi-ethnic evaluation teams to increase the chances of really hearing the voices of underrepresented students" (Stevens, 2000). Additionally, because there were a significant number of participants from various cultural backgrounds, data analysis techniques that examine differences in race and ethnicity were also appropriate for use in the program's context. Thus, racial and ethnic differences were found across the various populations in the evaluation, and this was evident in the evaluation. Further, the context of the evaluation program served to increase the success of the evaluation.

Implications of Findings

This study supports the use of culturally responsive evaluation. Presently, there are various resources in the evaluation field on culturally responsive evaluation. The two most notable are guidebooks released by the NSF. The sections on culturally responsive evaluation were utilized in this study to present the evaluation results. In addition to guidebooks on culturally responsive evaluation, there are also a handful of articles in the field that offer examples and lessons learned. Within this context, this study builds upon the current literature on culturally responsive evaluation practices and offers the field one additional example in the form of a case study to demonstrate both the application of culturally responsive evaluation throughout all phases of the evaluation, as well as a reflection of the benefits and challenges of this approach.

There are notable implications to be drawn from this study on the practice of culturally responsive evaluation. More specifically, the following recommendations are provided when evaluators attempt to be culturally responsive: 1) provide opportunities for stakeholders to become engaged in the evaluation process; 2) select evaluators who are from culturally diverse backgrounds and/or who represent the cultural backgrounds of the participants; 3) spend a significant amount time in the program context learning about the culture of the program, participants, and organization; 4) use multiple evaluation instruments that are sensitive to the cultural backgrounds of the participants; and 5) analyze and present data through a cultural lens.

Evaluators who seek to implement culturally responsive evaluations should take caution of a few items. First, evaluators should be aware that some evaluation contexts will not seem directly applicable. For example, clear dimensions of culture will not be present (i.e. the program does not serve underrepresented populations). However, culture is an integral part of any program and is present in any evaluation of a program, project, organization, and institution. Therefore, the concept of culture should be accounted for in every evaluation context and incorporated into the evaluation. Evaluations that do not take culture into account put the evaluation's validity, credibility, and use at risk. Further, "Failure to understand how cultural context interacts with program implementation and impact jeopardizes the validity of the evaluation" (Nelson-Barber, et al., 2005, p. 61).

Secondly, evaluators who seek to be culturally responsive should view culture to encompass more than race/ethnicity alone. While culture can be defined through ethnicity/race, culture also includes values, traditions, customs, history, socio-economic status, and language (Manswell-Butty, et al., 2004; Kirthart, 2005). Further, culture is transmitted from one generation to the next and includes learned and shared behaviors that influence the norms,

customs, spiritual and religious practices, language, educational systems, and other forms of living common to particular groups of society. Further, culture also includes the ways in which people identify with "collective groups such as organizations, institutions, communities, societies, and nations" (Kirthart, 2010, p. 401) Therefore, culture is more than the race/ethnicity of an individual, but also includes other factors which influence how individuals live, interact, and behave. Therefore, evaluators who adopt culturally responsive evaluation should respond to all facets of the culture of the participants.

The incorporation of culture beyond race/ethnicity can increase the quality of the evaluation. When culture is accounted for, than more than race/ethnicity, the evaluation can benefit from more accurate data collection and analysis. For example, responsiveness to culture in the data analysis phase can render a more complete picture of program outcomes. For example, disaggregation of data by race/ethnicity, socio-economic status, and language can reveal cultural differences in the participants' experiences (Frierson et al., 2010) Furthermore, qualitative data analyzed by culturally diverse evaluators can produce a more accurate portrayal of the participants' experiences in the program because they can relate and are familiar with the culture of the participants. When, and if possible, data should be analyzed in multiple and culturally sensitive ways in culturally responsive evaluation.

In addition to individual culture, evaluators who employ culturally responsive evaluation should also acknowledge that culture is also present at the organizational level (Kirkhart, 2010). For example, culture not only shapes the shared behaviors, attitudes, and beliefs of a group of people, but also impacts how organizations operate and interact. In this manner, "organizations, communities, and institutions impart cultural knowledge, values, beliefs, and skills that are shared and communicated across cohorts or generations (p. 401). Further, organizations are

situated within a specific set of circumstances, environmental, economic, and political (Thomas, 2004). In addition to culture at the individual level, evaluators who employ culturally responsive evaluation should not only respond to the culture of the participants, but also the culture of the program and the organization.

Thirdly, evaluators who employ culturally responsive evaluation should be cognizant that culture is fluid and ever-changing (Kirkhart, 2005). Therefore, the evaluator should take into account the unique cultural dimensions of each program that is evaluated because each one is different in scope, purpose, and structure. In addition, evaluators who aim to be culturally responsive should suspend all assumptions and expectations they have about particular groups in society. For example, evaluators should get to know the culture of the participants as is. The values, attitudes, behaviors, and facets of culture of the groups evaluated will vary beyond race. Differences in socio-economic status, gender, geographic location, etc. will impact the culture of the participants and the program. Therefore, the cultural dimensions will be dependent on the unique characteristics relevant to each program and its participants. Further, the cultural context of each program should be addressed adequately and appropriately for any given context.

Lastly, evaluators who employ culturally responsive evaluation should acknowledge that a certain level of cultural competence is necessary for implementing culturally responsive evaluation (Frazier-Anderson, et al., 2012). For example, while cultural competence refers to a set of skills that enable an evaluator to be competent in the culture of the evaluation, culturally responsive evaluation requires more than competency, but the direct inclusion of culture into the evaluation. Therefore, cultural competency and culturally responsive evaluation overlap in many ways; however there are clear differences. For example, cultural competency can refer to "what an evaluator knows," and "culturally responsive evaluation" can refer to "what an evaluator

does." Therefore, while an evaluator can argue that they are "cultural competent" in the culture of the evaluation, it does not necessary guarantee that they are culturally responsive. The next section discusses recommendations for future research.

Future Research

More practical guidelines. Future research on culturally responsive evaluation should investigate the ways in which culture and context can be further incorporated into the evaluation. More specifically, future research can identify additional strategies for conducting culturally responsive evaluation; these strategies should be applicable to more than one context. While the guidebooks on culturally responsive evaluation provide a number of examples of strategies that can be used to conduct culturally responsive evaluation, there is a need for research to further identify additional strategies for culturally responsive evaluation. Presently, the two guidebooks serve as a starting point for evaluators who would like to conduct culturally responsive evaluation.

More examples and literature. Additionally, future research should further examine the benefits and challenges of culturally responsive evaluation. To date, there are only a few examples in the field which provides examples of culturally responsive evaluation and highlight both the benefits and challenges of implementing this evaluation approach. However, more research and examples on culturally responsive evaluation are needed in order to develop a more complete understanding of the full spectrum of benefits and challenges associated with culturally responsive evaluation. Particularly, more research is needed on the benefits of the implementation of culturally responsive evaluation on programs and participants. For example, future studies on culturally responsive evaluation can highlight the benefits programs experience from the use of this type of evaluation, which can involve improved program quality and better

student experiences. While some researchers have already taken the liberty of discussing the challenges and benefits they faced when implementing this type of evaluation approach, further identification of all potential benefits of culturally responsive evaluation will help build further momentum for the use of culturally responsive evaluation and other evaluation approaches that take culture into account in the evaluation.

Need for quality criteria. Additionally, future research is also needed to focus on developing criteria for judging the quality of culturally responsive evaluation. While there are presently only a limited number of examples on the practice of culturally responsive evaluation, in the future, evaluators will have to judge the quality and extent to which evaluation efforts are culturally responsive. Criteria for judging the quality of culturally responsive evaluation should focus on the extent to which the evaluation fully describes and explains the cultural context of the program. Additionally, it should focus on the extent to which the data collected answered the evaluation questions and to what extent the evaluation data was valid. Further, the evaluation should examine outcomes based on the extent to which culture impacts the program context. The NSF guidebooks for conducting culturally responsive evaluation are a good starting point for developing such criteria. As more and more evaluators become attuned and familiar with the potential benefits of culturally responsive evaluation, more culturally responsive evaluations will need to be assessed as to what extent they are culturally responsive, and by what means.

Need for critical pool of minority evaluators. An important component of culturally responsive evaluation, which is less often addressed in the field of evaluation, is the lack of minority evaluators in the field of evaluation. In culturally responsive evaluation, evaluations are designed to be culturally responsive through the use of diverse evaluation team members. In culturally responsive evaluation, evaluators who share the same racial/ethnic background of the

participants can better understand the cultural context of the program and culture of the participants. In addition, they can better accurately collect and interpret the evaluation data and provide more accurate value judgments about the quality of the program (Fierson et al., 2002; Frierson et al., 2010). Further, "racial minority evaluators increase the democratic capacity and potential of institutions to the extent that they draw on their experiences and those of other marginalized communities (Prado, 2011, p. 418).

The type of evaluator is critical important in culturally responsive evaluation; however, there is clear evidence in the field that there are not enough evaluators who represent the cultural diversity present in the evaluation contexts in which they work. In addition, there is a clear lack of racial/ethnic minority evaluators who represent the disadvantaged racial/ethnic and socio-economic groups. To increase the number of minority evaluators, future research efforts should focus on developing methods and tools that non-minority or White evaluators can employ to develop some level of cultural competency in the culture in which evaluators work. In addition, research efforts should also develop tools and methods to guide evaluators in the development and implementation of culturally responsive evaluation. New and additional methods can enhance the current body of work on culturally responsive evaluation practice (i.e., NSF guidebooks). In addition, national professional organizations, for example, the AEA, should provide evaluators with workshops, trainings, and seminars on the theory, as well as the practical implications of culturally responsive evaluation. These workshops or training can also provide guidebooks and/or offer strategies that can guide evaluators in the evaluation process of culturally responsive evaluation.

More models. Future research should focus on the development of culturally specific evaluation models. Recently, a model of an African American Culturally Responsive Evaluation

was developed by African American scholars in the field (Frazier-Anderson, et al., 2012). The African American Culturally Responsive Evaluation System for Academic Settings (ACESAS) "is a logic model proposed to visually present the key steps used when implementing culturally responsive evaluation in majority African American communities with majority African American populations" (p. 347). In addition, the Indigenous Evaluation Framework (IEF) is a "culturally specific framework that both guides Indigenous evaluation and honors the importance of sovereign definitions of theory by individual's tribes. IEF position evaluations as a pathway to learning for the benefit of the tribal community" (Kirkhart, 2010, p. 406). Additionally, the TD Evaluation Framework (Thomas & Steven, 2004) is an evaluation framework designed for evaluating a specific set of programs: urban school reform interventions.

The frameworks above offer both conceptual and practical models for conducting different types of culturally responsive evaluations. The frameworks are helpful because they aid evaluation efforts; they also increase our general knowledge base on how to incorporate such concepts of culture and context into evaluation theory and practice. Additional models are needed in the field, both general and specific to provide guidance to evaluators who seek to employ culturally responsive evaluation. Further, culturally specific evaluation frameworks are needed for evaluating programs for conducting evaluations that serve large populations of Latina/o and Asian-American populations. These frameworks should be comprehensive in nature and include the various characteristics of these cultures. The concept of language will need to be addressed within these frameworks as language is a distinguishing characteristic of these cultures.

Conclusion

In more than one way, this study of culturally responsive evaluation provided the means to better understand the practice of culturally responsive evaluation. It demonstrated how a team of evaluators working from a culturally responsive evaluation approach planned for, designed, and implemented an evaluation that was inclusive of, and sensitive to, the cultural context of the program it evaluated. Moreover, the study highlighted examples of strategies that were developed and implemented throughout the evaluation process to conduct a culturally responsive evaluation. The strategies used were aligned with the strategies proposed in the two NSF Guidebooks on culturally responsive evaluation. Further, the evaluation met the guidelines set forth by the two guidebooks. In addition, the evaluation also employed strategies implemented in other culturally responsive evaluations. Therefore, the evaluation's use of culturally responsive evaluation reflected common practices in the literature on the practice of culturally responsive evaluation.

The G.A.M.E.S. program evaluation serves as an exemplar of culturally responsive evaluation. In culturally responsive evaluation, the success of an evaluation is dependent on the extent to which it has answered the evaluation questions (Frierson, et al., 2002). Further, the evaluation questions "themselves are always keys to a good evaluation" (p. 72). Based on the criteria above, the G.A.M.E.S. evaluation was a successful culturally responsive evaluation. The evaluation not only met its objectives, but also fully addressed the evaluation questions. In addition, program stakeholders were fully engaged throughout the evaluation, which ultimately increased the credibility and evaluation. Program improvements were evidence from year and year, and the experiences of the minority participants in the program were improved based on the evaluation results. In addition, data rendered in the evaluation was also valid and reliable based on the use of culturally sensitive evaluation instruments and data analysis. Further, the evaluation

results were disseminated widely and were relevant to all evaluation stakeholders (i.e., program staff, participants, administrators, etc.).

In conclusion, culturally responsive evaluation is a viable evaluation approach that has direct applicability to evaluation contexts. In the future, the evaluation field will need to move away from the assumption that culturally responsive evaluation is specific only to certain types of evaluation contexts (i.e., evaluations of programs in culturally diverse settings.) Evaluations are influenced by social, cultural, historical, and economic and political factors. In addition, evaluations are influenced by the evaluator's beliefs and predispositions (SenGupta, et al., 2004). Therefore, "Evaluations are thus far from being value-free and culture-free (SenGupta, Hopson, & Thompson-Robinson, 2004), as culture and values permeate all facets of social programs and their evaluations" (Chouinard & Cousins, 2010, p. 457). In defense of culturally responsive evaluation, Frierson et al. (2010) argue:

> To ignore the reality of the influence of culture and to be unresponsive to the needs of the target population puts the program in danger of being ineffective and the evaluation in danger of being seriously flawed. Evaluation should serve the public good by presenting valid information about programs that have been properly evaluated. (p. 76)

Further, culturally responsive evaluation represents plain "good evaluation" (Hood, 2011). The components of culturally responsive evaluation are the same components that have proven successful for other evaluation approaches or designs (i.e., goals-based evaluation and utilization-focused evaluation). For example, the use of mixed methods has proven success in evaluation that has employed various types of evaluation approaches that have sought to capture the richness and complexity of these programs. Therefore, culturally responsive evaluation is "good evaluation and it permeates all phases of the evaluation enterprise" (NSF Project Evaluation Guide, 2010). Therefore, culturally responsive evaluation is a successful evaluation

approach that can aid in the success of an evaluation by providing valid, credible, and useful valid evaluation results to program stakeholders.

References

Abma, T. A. & Stake, R. E. (2001). Stake's responsive evaluation: Core ideas and evolution. *New Directions for Evaluation, 92,* 7-21.

American Evaluation Association (2004). Guiding principles for evaluators. Retrieved from http://www.eval.org/publications/GuidingPrinciplesPrintable.asp

American Evaluation Association (2011). The public Statement on Cultural Competence in Evaluation. Retrieved from http://www.eval.org/ballot/ccstatementvote.asp

Askew, K., Beverly, M. G., & Jay, M. L. (in press). Aligning collaborative and culturally responsive evaluation approaches. *Evaluation and Program Planning.* Doi: 10.1016/j.evalprogplan.2011.12011

Blickenstaff, J.C. (2005). Women and science careers: Leaky pipeline or gender filter?, *Gender and Education, 17*(4), 369-386.

Botcheva, L., Shih, J., & Huffman, L. C. (2009). Emphasizing Cultural Competence in Evaluation: A process-oriented approach. American Journal of Evaluation, 30(2), 176-188.

Burke, J.R., & Mattis, M.C. (Eds.). (2007). *Women and minorities in science, technology, engineering, and mathematics: Upping the numbers*. Massachusetts: Edward Elgar Publishing.

Butty, J. L., Ried, M., & LaPoint, V. (2004). A culturally responsive evaluation approach applied to the Talent Development School-to-Career intervention program. *101*, 37-47.

Chouinard, J. A. & Cousins, J.B. (2009). A review and synthesis of current research on cross-cultural evaluation. *American Journal of Evaluation, 30*(4), 457-494.

Clewell, B. C., & Anderson, B. (1991). *Women of color in mathematics, science and engineering: A review of literature*. Washington, DC: Center for Women Policy Studies.

Clewell, B.C., Anderson, B.T., & Thorpe, M. (1992). *Breaking the barriers: Helping female and minority students succeed in mathematics and science.* San Francisco: Jossey-Bass.

Clewell, B. C., & Campbell, P.G. (2002). Taking stock: Where we've been, where we are, where we're going. *Journal of Women and Minorities in Science and Engineering, 8,* 255-284.

Conner, R.F. (2004). Developing and implementing culturally competent evaluation: A discussion of multicultural validity in two HIV prevention programs for Latinos. *New Directions for Evaluation, 102,* 51-65.

Cousins, J. B. & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation, 80,* 5-23.

Denzin, N. (1978). *The research act*. New York: McGraw-Hill.

Fetterman, D. M. (1994). Empowerment Evaluation. *Evaluation Practice*, *15*(1), 1-15.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2003). *Program evaluation: Alternative approaches and practical guidelines*. (3rd ed.). Boston: Pearson Education.

Frazier-Anderson, P., Hood, S. & Hopson, R.K. (2011). Preliminary considerations of an African American culturally responsive evaluation system. In S.D. Lapan, M.T. Quartaroli, & F.J. Riemer (Eds.), *Qualitative Research: An Introduction to Methods and Designs*, San Francisco: Jossey-Bass.

Frierson, H. T., Hood, S., & Hughes, G. (2002). Strategies that address culturally responsive evaluation. In J. Frechtling (Ed.), *The 2002 user friendly handbook for project evaluation*. Arlington, VA: National Science Foundation.

Frierson, H. T., Hood, S., Hughes, G. B., and Thomas, V. G. (2010). A guide to conducting culturally responsive evaluations. In J. Frechtling (Ed.), *The 2010 user friendly handbook for project evaluation.* Arlington, VA: National Science Foundation.

Greene, J., Lincoln, Y., Mathison, S., Mertens, D. M., & Ryan, K. (1998). Advantages and challenges of using inclusive evaluation approaches in evaluation practice. *American Journal of Evaluation, 1*, 101-102.

Greene, J. C. (2000). Challenges in practicing deliberative democratic practice. *New Directions for Evaluation, 85,* 13-26.

Greene, J. C. (2005). Evaluators as stewards of the public good. In Hood, S. Hopson, R. H., & Frierson, H. (Eds.), *The role of culture and cultural context: A mandate for inclusion, the discovery of truth, and understanding in evaluative theory and practice*, (pp. 87-104). Greenwich, CT: Information Age.

Greene, J.C. (2007). Mixed methods in social inquiry. San Francisco: Jossey-Bass.

Greene, J. C., DeStefano, L., Burgon, H., & Hall, J. (2006). An educative, values-engaged approach to evaluating STEM educational programs. *New Directions for Evaluation*, 109, 53-71.

Guba, Y.S. & Lincoln, E.G. (1985). *Naturalistic Inquiry*. Newbury Park: CA: Sage.

Guzman, B. L. (2003). Examining the Role of Cultural Competency in Program Evaluation: Visions for New Millennium Evaluators. In S. I. Donaldson and M. Scriven (Eds.),

Evaluating Social Programs and Problems: Visions for the New Millennium, (pp. 167-181). Mahwah, N.J: Lawrence Erlbaum Associates.

Hartwig, B. (2010). Understanding Cultural Nuances Creates Global Success. *Diversity Inc Magazine.* Retrieved from http://www.diversityinc.com/content/1757/article/6283/

Hilton, G. (2007). Becoming culturally fluent: understanding the variances of culture across Asia can enhance business and communication. *Communication World, Nov-Dec.* Retrieved from http://www.entrepreneur.com/tradejournals/article/171659325.html

Hood, S. (2000). Commentary on deliberative democratic evaluation. *New Directions for Evaluation, 85,* (77-83).

Hood, D. & Cassaro, D. A. (2002). Feminist evaluation and the inclusion of difference. *New Directions for Evaluation, 96,* 27-40.

Hood, S., Hopson, R., & Frierson, T. (2005). *The role of culture and cultural context in evaluation: A mandate for inclusion, the discovery of truth and understanding*. New York: NY: Information Age.

Hood, S., & Hopson, R. K. (2008). Evaluation roots reconsidered: Asa Hilliard, a fallen hero in the Nobody knows project, and African educational excellence. *Review of Educational Research, 78,* 410-426.

Hood, S. (May, 2011). Personal communication.

Hopson, R. (2009). Reclaiming knowledge at the margins: Culturally responsive evaluation in the current evaluation moment. In K. E. Ryan & J. B. Cousins (Eds.), *The Sage International handbook of educational evaluation* (pp. 429-446). Thousand Oaks, CA: Sage.

Hopson, R. (2000). Editor's Notes. How and why language matters in evaluation. *New Directions for Evaluation*, *86,* 1-3.

Hopson, R. (2010, March, 7). Rodney Hopson on Culturally Responsive Evaluation. AEA 365: Tip-a-Day by and for Evaluators [Web log post]. Retrieved from http://aea365.org/blog/?p=247

Hopson, R. K., Lucas, K. J., & Peterson, J.A. (2000). HIV/AIDS talk: Implications for prevention intervention and evaluation. *New Directions for Evaluation, 86,* 29-42.

House, E. R. & Howe, K. R. (1999). *Values in evaluation and social research and social research*. Thousand Oaks, CA: Sage.

House, E. R. (1999). Evaluation and people of color – A response to Professor Stanfield. *American Journal of Evaluation, 20(3),* 433-435.

House, E. R. & Howe, K. R. (2000). Deliberative democratic evaluation. *New Directions for Evaluation, 85,* 3-12.

House, E. R. (2001). Responsive evaluation (and its influence on deliberative democratic evaluation). *New Directions for Evaluation, 92,* 23-30.

Howe, K. & Ashcraft, C. (2005). Deliberative democratic evaluation: Successes and limitations of an evaluation of school chose, *Teachers College Record, 107*(10), 2274-2297.

International Technology Education Association (2002). *Standards for Technological Literacy: Content for the Study of Technology*. Retried January 1, 2012, from http://www.iteaconnect.org/TAA/PDFs/xstnd.pdf

King, J.A. (1998). Making sense of participatory evaluation. *New Directions for Evaluation, 108,* 57-67.

King, J.A. (2007). Making sense of participatory evaluation. *New Directions for Evaluation, 114,* 83-104.

Kirkhart, K. E. (2005). Through a cultural lens: Reflections on validity and theory in evaluation. In Hood, S. Hopson, R., & Frierson, H. (Eds.), *The role of culture and cultural context in evaluation: A Mandate for inclusion, the discovery of truth, and understanding in evaluative theory and practice,* (pp. 87-104). Greenwich, CT: Information Age.

Kirkhart, K. E. (2010). Eyes on the prize: Multicultural validity and evaluation theory. *American Journal of Evaluation, 31,* 400-413. 83.

Ladson-Billings, G., & Tate, W. F. (1995). Toward a critical race theory of education. *Teachers College Record, 97,* 47-68.

LaPoint, V. & Jackson, H. L. (2004). Evaluating the co-construction of the Family, School, and Community Partnership in a low-income urban high school. *New Directions for Evaluation, 101,* 25-36.

Madison, A. M. (2000). Language in defining social problems and evaluating social programs. *New Directions for Evaluation, 86,* 17-28.

Manswell-Butty, J.L., Reid, M.D., & LaPoint, V. (2004). A culturally responsive evaluation approach applied to the Talent Development Model School-to-Career Intervention program. *New Directions for Evaluation, 101,* 37-47.

McDonald, D.A., Kutura, P.B.C., Richmond, L.S., & Betts, S.C. (2004). Culturally Respectful Evaluation. *The Forum for Family and Consumer Issues* (FFCI), *9*(3), Retrieved April 1, 2012, from http://ncsu.edu/ffci/publications/2004/v9-n3-2004-december/ar-1-cuturally.php.

Mertens, D. M. (1999). Inclusive evaluation: Implications of transformative theory for evaluation. *American Journal of Evaluation*, *20*(1), 1-14.

Mertens, D. M. (2003). The Inclusive View of Evaluation: Visions of the new Millennium. In S. I. Donaldson and M. Scriven (eds.), *Evaluating Social Programs and Problems: Visions for the New Millennium*, 91-107. Mahwah, N.J.: Erlbaum.

Mertens, D.M. & Hopson, R.K. (2006). Advancing evaluation of STEM efforts through attention to diversity and culture. *New Directions for Program Evaluation*, *109*, pp. 35-51.

National Action Council for Minorities in Engineering Incorporation. (2008). *Confronting the "New" American Dilemma: Underrepresented Minorities in Engineering: A Data-Based Look at Diversity*. Retrieved from http://www.nacme.org/user/docs/NACME%2008%20ResearchReport.pdf

National Research Council (1996). *National Science Education Standards.* Washington, DC: The National Academies Press. Retrieved Jan 1, 2012, from http://www.nap.edu/openbook.php?record_id=4962&page=1

National Science Foundation. *Online Evaluation Resource Library*. Retrieved January 1, 2012, from http://oerl.sri.com/home.html

Nelson-Barber, S., LaFrance, J., Trumbell, E., & Aburto, S. (2005). Promoting culturally reliable and valid evaluation practice. In S. Hood, R.K. Hopson, & H. Frierson (Eds.), *The role of culture and cultural context in evaluation* (pp. 61-85). Greenwich, CT: Information Age.

Oakes, J. (1990). Lost Talent: The underrepresentation of women, minorities, and disabled persons in science (Report No. R-3774-NSF/RC). Santa Monica, CA: National Science Foundation/Rand Corporation.

Oakes, J., Ormset, T., Bell, R., & Camp, P. (1990). Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science (Report No. R-3928-NSF). Santa Monica, CA: National Science Foundation/Rand Corporation.

Page, J.B. (2005). The concept of culture: A core issue in health disparities. *Journal of Urban Health, 82* (3), iii35-iii43.

Parker, L. (2004). Can critical theories of or on race be used in evaluation research in education? *New Directions for Evaluation, 101,* 85-93.

Patton, M.Q. (2000). Overview: Language matters. *New Directions for Evaluation, 86,* 5-16.

Pewewardy, C.D. (1994). Culturally responsive pedagogy in action: An American Indian magnet school. In E.R. Hollins, J.E. King, & W.C. Hayman (Eds.), *Teaching Diverse*

Populations: Formulating a Knowledge Base (pp. 77-92). Albany: State University of New York Press.

Philips, S.U. (1983). *The Invisible Culture: Communication in Classroom and Community on the Warm Springs Indian Reservation*. Prospect Heights, IL: Waveland.

Pillow, W. S. (2002). Gender matters: Feminist Research in Educational Evaluation. *New Directions for Evaluation, 96,* 9-26.

Prado, J. (2011). "Honor the Context": Opening Lines for a Critical Multicultural Evaluative Practice. *American Journal of Evaluation, 32*(3), 418-427.

Ryan, K. E., Chandler, M., & Samuels, M. (2007). What should school-based evaluation look like? *Studies in Educational Evaluation, 33,* 197-212.

Samuels, M. & Ryan, K. (2011). Grounding evaluation in culture. *American Journal of Evaluation, 32*, 183- 198.

SenGupta, S., Hopson, R. & Thompson-Robinson, M. (2004). Cultural competence in evaluation: An overview. *New Directions for Evaluation, 102*, 5-19.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects, *Education for Information, 22*, 63–75.

Smith, L. T. (1999). *Decolonizing methodologies: Research and indigenous peoples*. London, UK: Zed.

Stake, R. E. (1975). Program evaluation, particularly responsive evaluation. *Occasional Paper No. 5* Retrieved March 1, 2011, from http://www.wmich.edu/evalctr/pubs/ops/

Stake, R. E. (1995). *The Art of Case Study Research*. Thousand Oaks: Sage Publications.

Stake, R.E. (2003). *Standards-based and responsive evaluation.* Thousand Oaks, CA: Sage.

Steele, C.M., & Aronson, J. (1995). Stereotype treat and the intellectual performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797-811.

Steele, C.M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*(6), 613-629.

Symonette, H. (2004). Walking pathways toward becoming a culturally competent evaluator: Boundaries, borderlands, and border crossings. *New Directions for Evaluation, 102*, 95-109.

Thomas, V. G. (2004). Building a contextually responsive evaluation framework: Lessons from working with urban school interventions. *New Directions for Evaluation, 101,* 3-24.

Thomas, V. G. & Stevens, F. I. (2004). Editor's notes. *New Directions for Evaluation, 101,* 1-2.

University of Illinois at Urbana-Champaign. *Girls' Adventures in Mathematics, Engineering, and Science Summer Camp*. Retrieved April 1, 2011, from https://wiki.engr.illinois.edu/display/games/G.A.M.E.S.+Camp

United States Department of Commerce. (2011). *Education Supports Racial and Ethnic Equality in STEM*. Retrieved Dec 1, 2011 from http://www.nacme.org/user/docs/NACME%2008%20ResearchReport.pdf

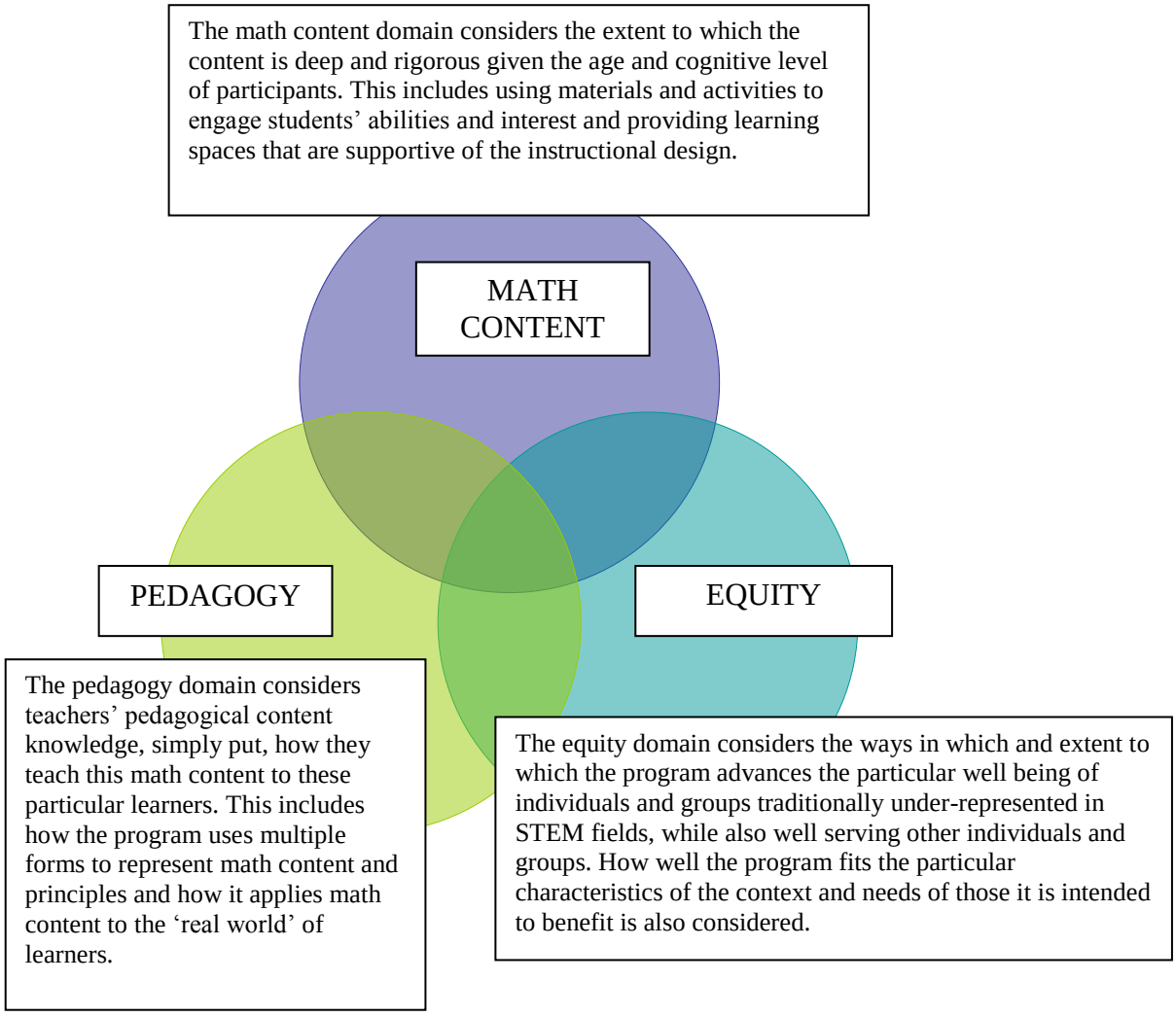
United States Department of Labor. (2007). *The STEM Workforce Challenge: the role of the Public Workforce System in a National Solution for a Competitive Science, Technology, Engineering, and Mathematics (STEM) Workforce*. Retrieved December 1, 2011, from http://www.doleta.gov/youth_services/pdf/STEM_Report_4%2007.pdf

Yin, R. (1994). *Case study research: Design and methods* (2nd ed.). Beverly Hills, CA: Sage Publishing.

Zulli, R. A., & Frierson, H. T. (2004). A focus on cultural variables in evaluating an Upward Bound Program. *New Directions for Evaluation, 102,* 81-93.

Appendix A
The Educative, Values-Engaged Framework

The following diagram presents the three key domains of the educative, values-engaged approach to evaluation. Within each domain, some guiding considerations are offered.



From Hall, J., Greene, J., DeStefano, L., & Johnson, J., (2006, November). Ready, Set-Evaluate! Field Study Results from an Educative, Values-Engaged Evaluation of Science, Technology, Engineering, and Mathematics (STEM) Educational Programs. Roundtable discussion presented at the annual meeting of the American Evaluation Association, Portland, OR.

Appendix B

Evaluation Presentation Slides

Percentage of Campers Who Agreed with the Following Statements about Lessons.

	Structures	Computer Science	Bio/Chem. Engineering	BioImaging
Most lessons were too difficult	4	4	12	17
A few of the lessons were too difficult.	12	31	21	59
Most lessons were just right.	67	88	79	83
A few of the lessons were too easy.	35	34	40	29
Most lessons were too	31	13	21	8

Structures – Camper Comments

"All classes were just right; I enjoyed them a lot!"

"My team and I worked hard to build a boat and with teamwork we succeeded in doing that."

"I thought that the classes were taught well so for me everything was just right."

Computer Science – Camper Comments

"*Squeak* was a challenge to pick up at first, but after a couple of hours I started to accomplish the tasks without making any mistakes."

"Our group was awesome, we got along really well. My project group was very good because we worked well together to make a good project."

Bio/Chem. Engineering– Camper Comments

"It was very fun except sometimes things were a little rushed. Also, there were few too many slideshows. I really like the Olympics games though."

"I learned how to cooperate with all kinds of people and they helped me think of new ideas I never would have known about."

BioImaging – Campers' Comments

"I thought that the talks and explanations went really fast and at times it was a bit hard to follow. I think it was pretty hard to keep track of time since we had to do so much during this camp! Also, the talks could have been a little more interesting. Overall, I definitely had fun, and I learned a lot of new information."

How does G.A.M.E.S. influence attitudes towards STEM?

- Most campers started the camp already interested in engineering and slightly more likely to pursue a career in engineering.
- Campers' perceptions about people interested in STEM changed after participating in this camp.

Appendix C

Observation Guide

The main purposes of the classrooms observations are to:

1. To provide a descriptive record of the pedagogy, content, and classroom/lab activities.
2. To capture a descriptive record of students' participation in the classes/lab, noting the students' attention, engagement with daily activities, enthusiasm as well as boredom, disinterest and/or frustration. Special attention should be paid to students' interactions with instructors, and each other in group projects.
3. To observe ways in which the program's content and pedagogy are aligned, or mis-aligned with the students' particular learning style.

Description of context, setting:

1. Describe the context / setting, including its physical features, how teacher and students are arranged in the room, and the general classroom atmosphere, using as much descriptive language as possible. A sketch of the physical setting and how people are arranged is also encouraged.

Descriptive account of program activities:

1. Document the key activities (e.g., presentations, group activities, video, lectures, hands-on, paper / pencil tasks, or outdoor activities). Provide a general description of the activity. For a piece of the class or the entire class period, record each activity; and indicate how long each one took.
2. For each activity, provide a description of the learners' attention, engagement, enthusiasm, as well as boredom, disinterest, frustration (in a few sentences).
3. Provide a general description of the instruction (e.g., interactive, directive, amount of response time provided to learners, pace of instruction, use of learners' prior knowledge, questioning, recall, incentives, etc…), use of materials (e.g., visuals, manipulatives), and engagement with students (e.g., encouraging, distant, responsive, etc…) as appropriate. (again, in a few sentences)
4. Each write-up may vary from 2 to 7 pages per observation.

Descriptive account of learners' participation:

1. For each observation, randomly select 4-5program participants for a focused observation, to ensure a large sample of participants will have been observed after multiple observations.

2. This focused observation may include learners who sit near each other or engage in small group activities.
3. Observe the students in this sub-group for approximately 15 minutes during the class.
4. If the opportunity arises, one observation should be made during an entire class activity and another should be made during a small group or individual activity.
5. If there is an unexpected event or response that occurs during class, this should be descriptively noted as well.
6. Focused observations should target the following: (a) the character of the learners' attention, engagement, enthusiasm, as well as frustration, boredom, or disinterest; (b) students' interactions with one another, the instructor, and the materials; and (c) any other event or response that relates to the students' experience.

Appendix D

Evaluator Interview Protocol

General Questions Relating to Culturally Responsive Evaluation (CRE)

1. How would you define and/or describe CRE?
2. How would you describe the major goals and/or aims of CRE?
3. In your opinion, what are most significant aspects of CRE?
4. Do you have previous evaluation experiences in which you have applied CRE? If yes, please describe the three most significant experiences that stand out as being CRE?
    a. What were the strengths of these evaluations in terms of CRE?
    b. What could have been improved upon in terms of CRE? (i.e. limitations)

General Questions Relating to G.A.M.E.S. and CRE

1. What was the purpose of the evaluation?
2. Please describe the evaluation context (factors that were present within the evaluation that influenced program results such as geographic location, timing, political and social climate, and economic conditions, relationships between people, political dynamics, etc.).
3. From a CRE perspective:
    a. How did you prepare for the evaluation?
        i. The evaluation team was diverse (from different racial/ethnic backgrounds)
        ii. The evaluation team had an understanding of the cultural context of the program
        iii. The evaluation team had an understanding of the culture of the participants.
    b. What factors did you consider when designing the evaluation?
        i. The evaluation team addressed questions of significance in the evaluation
        ii. The evaluation team addressed appropriate questions in the evaluation
        iii. The evaluation questions were carefully considered by the program staff
        iv. The evaluation team articulated what would be accepted as evidence when to the evaluation questions
        v. The evaluation design was robust, had qualitative and quantitative components

vi. Instruments were 1) valid, 2) reliable, and 3) culturally responsive
vii. Evaluation team used qualitative evaluation techniques (via interviews, observations, etc.)

c. What factors did you consider when implementing the evaluation?

i. The evaluation team engaged stakeholders (i.e. took into account stakeholders' perspectives and views from a cultural perspective)
ii. The evaluation team took into account stakeholders perspectives in the interpretations and presentation of the results
iii. The evaluation team analyzed the data from a culturally responsive evaluation perspective (i.e. accurate meaning of data was made)
iv. The evaluation team disseminated the evaluation results broadly (via presentations or written reports)
v. The evaluation results were useful

4. To what extent did the evaluation meet your expectations in terms of both the quality of the evaluation and the outcomes of the evaluation?

5. What were some challenges you experienced during the evaluation? How were these challenges addressed?

6. What were some of the strengths or highlights of the evaluation?

7. Can you think of other ways the evaluation team was culturally responsive?

8. Are there other techniques that you think could have contributed to a more culturally responsive evaluation?

9. Is there anything else you would like to add regarding the G.A.M.E.S. evaluation and its use of a CRE approach?

Appendix E

Program Administrator Interview Protocol

Questions Related to G.A.M.E.S. Evaluation

1. What were some of the reasons for requesting the evaluation?
2. How would you describe the purpose of the evaluation?
3. What are the kinds of evaluation activities that occurred? Of those which do you see as the important? Why were these most important?
4. What needs did you expect to be met by the evaluation? To what extent were they met? If not, what were the shortcomings of the evaluation?
5. When you first requested the evaluation, what were your expectations for the evaluation? What were your plans for the evaluation findings?
6. In retrospect, did the evaluation meet your expectations? (match between what you wanted and what you got)
7. What do you see as the three most important strengths of the evaluation?
8. What were some of the limitations of the evaluation?

Questions Related to Culturally Responsive Evaluation

9. Are you familiar with culturally responsive evaluation?
10. How would you define and/or describe CRE?
11. From the document review of the evaluation reports, it shows that the evaluation team attempted to be culturally responsive. From your perspective, what are examples of tools, strategies, or techniques in which the evaluation team used to design and/or implement a CRE?
12. More specifically, which of the following were evident?
    a. The evaluation team was diverse (from different racial/ethnic backgrounds)
    b. The evaluation team had an understanding of the cultural context of the program
    c. The evaluation team had an understanding of the culture of the participants.
    d. The evaluation team engaged stakeholders (i.e. took into account stakeholders' perspectives and views from a cultural perspective)
    e. The evaluation team took into account stakeholders perspectives in the interpretations and presentation of the results
    f. The evaluation team addressed questions of significance in the evaluation
    g. The evaluation team addressed appropriate questions in the evaluation
    h. The evaluations were carefully considered by the program staff
    i. Evaluation team and evaluators articulated what will be accepted as evidence when to the evaluation questions

j. The evaluation design was robust, had qualitative and quantitative components
k. Instruments were 1) valid, 2) reliable, and 3) culturally responsive
l. Evaluation team used qualitative evaluation techniques (via interviews, observations, etc.)
m. The evaluation team analyzed the data from a culturally responsive evaluation perspective (i.e. accurate meaning of data was made)
n. The evaluation team disseminated the evaluation results broadly (via presentations or written reports)
o. The evaluation results were useful

13. Can you think of other ways the evaluation team was culturally responsive?
14. Are there other techniques that you think could have contributed to a more culturally responsive evaluation?
15. Is there anything else you would like to add regarding the G.A.M.E.S. evaluation and its use of a CRE approach?

Appendix F

Rubric

Dimensions of CRE	0 Not at all	1 To some extent	2 To a moderate extent	3 To a great extent	Examples of presence
Identify evaluation stakeholders					
Provide a full description of the program context					
Develop evaluation questions that are important					
Describe in detail the evaluation activities					
Provide a clear description of the evaluation's purpose					
Design a rigorous evaluation plan with qualitative and quantitative methods					
Develop instruments that are culturally responsive					

Dimensions of CRE	0 Not at all	1 To some extent	2 To a moderate extent	3 To a great extent	Examples of presence
Triangulate data					
Analyze program impact on program participants					
Disaggregate Data					
Analyze program impact within the context of the program					
Report evaluation findings on the successes, challenges, and limitations of the program					
Provide recommendations that are clearly linked the evaluation findings					
Total Number:					

Appendix G

Sample Materials

Introduction to Aseptic Technique

Bacteria and fungi live all around us, and they are especially fond of the warm, moist conditions in which we grow cells. This means that contaminants which find their way into cell cultures will take over. These bacteria and fungi will steal away all the nutrients we put in to help the cells grow, generally leaving the cells to die. To prevent this, we have to be very careful not to let bacteria or fungi get anywhere near the area where we will work with the cells. Below are some of the things we do to reduce the likelihood of contamination. For your cells to grow safely, you need to be careful to follow all these suggestions. When we do all these things we say we are following **aseptic technique**.

- Wear short sleeves or roll sleeves up
- Tie back long hair
- Remove rings and watches
- Wash hands with soap and water
- Spray hands, the work surface, and equipment with 70% ethanol
- Do not breathe into cultures, bottles of media, etc.
- Keep talking to a minimum; no singing or chewing gum
- Keep bottles and flasks closed when you are not working with them
- Use only sterilized pipettes, plates, flasks, and bottles in the biosafety hood
- Change pipettes for each manipulation; if the tip touches something outside the flask, replace with a new one
- Do not let media slosh near the rims of flasks or plates

Lab Exercise: Fluorescent Staining of Actin

Supplies

One per group:

- Box of Kimwipes
- 1 piece of Parafilm with 2 coverslips coated with fibronectin and seeded with fixed cells
- 1 glass microscope slide
- Foil box
- 10 ml PBS
- 1 ml glycine in tube
- 1 ml Triton X-100 in tube
- Liquid waste container
- Micropipettor (1000 µl) with tips

For the room:

- 70% ethanol in spray bottle
- Paper towels
- Mounting media
- Forceps

What are all these weird things? Turn to page 38 for information about all the strange words we use so you can understand what we're talking about.

Lab Exercise: Fluorescent Staining of Actin

Procedure

1. Wash your hands.
2. On your lab bench are two coverslips, each a small circle of clear, colorless glass. The top side of the coverslip has cells fixed (attached) to it. These coverslips have been rinsed with PBS and glycine. The liquid glycine is sitting on the coverslips.
3. Aspirate (suck up) the glycine from each coverslip with a 1000 µl pipettor. Place the glycine in your "liquid waste" container. Using a new pipette tip, add fresh glycine to each coverslip drop wise — squirt the liquid out gently. Let sit for 5 minutes.
4. Aspirate the glycine as before. With a new pipette tip, rinse each coverslip 5 times with PBS. Remember to add the liquid drop wise.
5. Aspirate the last PBS wash and add Triton X-100. Let it stand for 10 minutes (set your timer).
6. Aspirate the Triton X-100 and rinse 5 times with PBS. Place a wet paper towel under the Parafilm.
7. Aspirate the last of the PBS. Ask a lab assistant to place 200 µl of rhodamine phalloidin on each of the coverslips. Cover both coverslips and the wet paper towel with the foil box. Let stand for 30 minutes (set your timer).
8. Ask the lab assistant to aspirate the rhodamine phalloidin. You should then rinse the coverslips 5 times with PBS.
9. Clean the glass microscope slide with 70% ethanol. Place 25 µl warm mounting media in two spots on the slide (ask a lab assistant for help if necessary).
10. Remove the last rinse of PBS from the coverslips. Ask a lab assistant to place your coverslips on the slide. Cover your slide with the foil box.
11. Keep your slide under the foil box at all times. During the next activity, you will view your cells, and the actin will be glowing red!

Appendix H

Sample Day Schedule for G.A.M.E.S.

7:00 AM-7:59 AM	Wake up
8:00 AM-8:50 AM	Breakfast
9:00 AM-11:50 AM	Class Sessions and hands-on demos
12:00 PM-12:50 PM	Lunch
1:00 PM-2:29 PM	Team Projects or afternoon activity
2:30 PM-3:59 PM	Special Speakers, engineering tours
4:00 PM-5:50 PM	Team Projects
6:00 PM-6:50 PM	Dinner
7:00 PM-9:10PM	Evening activities (swimming, African dancing, craft, other)
9:20 PM-10:00 PM	Getting Ready for bed
	Good Night- Lights OUT

Appendix I

Participant Survey

Print Name:

Please Check One:

1) I am interested in engineering.

☐ Strongly Agree ☐ Agree ☐ Not Sure ☐ Disagree ☐ Strongly Disagree

2) Students who are good in math and science get picked on, teased, or ignored by other students.

☐ Strongly Agree ☐ Agree ☐ Not Sure ☐ Disagree ☐ Strongly Disagree

3) I am interested in pursuing a career in engineering.

☐ Strongly Agree ☐ Agree ☐ Not Sure ☐ Disagree ☐ Strongly Disagree

4) People who enter math and science careers tend to be "nerds" who lack social skills.

☐ Strongly Agree ☐ Agree ☐ Not Sure ☐ Disagree ☐ Strongly Disagree

5) Scientists and engineers work longer hours than most other professionals.

☐ Strongly Agree ☐ Agree ☐ Not Sure ☐ Disagree ☐ Strongly Disagree

6) Having a career in engineering or science requires too many years of education.

☐ Strongly Agree ☐ Agree ☐ Not Sure ☐ Disagree ☐ Strongly Disagree

7) Becoming a scientist or engineer is a good way to serve humanity.

☐ Strongly Agree ☐ Agree ☐ Not Sure ☐ Disagree ☐ Strongly Disagree

8) I have a good understanding scientists and engineering actually do.

☐ Strongly Agree ☐ Agree ☐ Not Sure ☐ Disagree ☐ Strongly Disagree

9) Scientists and engineers mainly work alone, with equipment instead of people.

☐ Strongly Agree ☐ Agree ☐ Not Sure ☐ Disagree ☐ Strongly Disagree

10) Girls who are good in math and/or science are just as popular with boys as other girls are.

☐ Strongly Agree ☐ Agree ☐ Not Sure ☐ Disagree ☐ Strongly Disagree

11) I am confident that I can do science/engineering.

☐ Strongly Agree ☐ Agree ☐ Not Sure ☐ Disagree ☐ Strongly Disagree

12) I can picture myself working in science or engineering someday.

☐ Strongly Agree ☐ Agree ☐ Not Sure ☐ Disagree ☐ Strongly Disagree

13) Working in a science laboratory **would not** be an interesting way to earn a living.

☐ Strongly Agree ☐ Agree ☐ Not Sure ☐ Disagree ☐ Strongly Disagree

14) Do you know anyone that has a job that involves science, technology, engineering or technology?

☐ Yes ☐ No

15) If you could continue with your schooling as far as you wanted, what is the highest level of education you would complete? Check one.

☐ High School

☐ Community College (two-year college program) or Vocational School

☐ College (four or five year college program)

☐ Graduate School - Master's Degree

☐ Graduate School - Doctoral Degree (Ph.D.)

☐ Medical, Dental, or Veterinary School

☐ Law School

☐ Other (specify _____)

16) Realistically, what is the highest level of education you *expect* to complete? Check one.

☐ High School

☐ Community College (two-year college program) or Vocational School

☐ College (four or five year college program)

☐ Graduate School - Master's Degree

☐ Graduate School - Doctoral Degree (Ph.D.)

☐ Medical, Dental, or Veterinary School

☐ Law School

☐ Other (specify _____)

17) How likely is it that you will take the following courses?

	Very Likely	Somewhat Likely	Somewhat Unlikely	Very Unlikely	Already Taken
Algebra:	1	2	3	4	5
Geometry:	1	2	3	4	5

Trigonometry:	1	2	3	4	5
Pre-Calculus:	1	2	3	4	5
Calculus:	1	2	3	4	5
AP Calculus:	1	2	3	4	5
Biology:	1	2	3	4	5
AP Biology:	1	2	3	4	5
Chemistry:	1	2	3	4	5
AP Chemistry:	1	2	3	4	5
Physics:	1	2	3	4	5
AP Physics:	1	2	3	4	5

18) How would you rate most of the lessons you experienced during the G.A.M.E.S. camp?

	Strongly Agree	2	3	Strongly Disagree
Most of the lessons were too difficult.	1	2	3	4
A few of the lessons were too difficult.	1	2	3	4
Most of the lessons were just right.	1	2	3	4
Most of the lessons were too easy.	1	2	3	4
A few of the lessons were too easy.	1	2	3	4

19) Please provide comments about lessons that you thought were too easy or too difficult.

20) Please share your opinions about the Motorola presentation.

_____

_____

21) Please share your opinions about CRCE/ARC nights.

_____

_____

22) Please share your experiences with teamwork this week.

_____

_____

23) Directions:
Think about your participation in the polio vaccine project this week. Read each statement below and indicate what you did.

1) I helped to solve problems by using information provided by the team.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

2) I exercised leadership.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

3) I focused on completing the team task successfully.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

4) I worked well with girls from other races.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

5) I taught other team members.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

6) I identified possible alternatives.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

7) I attempted to change incorrect information immediately.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

8) I organized team activities to complete tasks on time.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

9) I used available information to make decisions.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

10) I interacted cooperatively with other team members.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

11) I clearly and accurately exchanged information with team members.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

12) I conducted myself with courtesy.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

13) I evaluated what would happen if the team made other choices.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

14) I tried to figure out what the team was thinking so as to solve problems.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

15) I showed that I understood the information.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

16) I served as a role model in for the other girls.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

17) I helped other team members.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

18) I understood and contributed to the team's goals.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

19) I paid attention to what others are saying.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

20) I respected the thoughts and opinions of others in the team.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

21) I kept track of time.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

22) I looked for a team approach.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

23) I led when appropriate, helping the group do very well.

☐ Almost Never ☐ Sometimes ☐ Often ☐ Almost Always

24) Is there anything else you would like for G.A.M.E.S. to know about your experience?

_____

_____

Bonus Questions

1) What are networks? Where do you find networks in the real world?

2) How do you count in binary? If you have the binary number '100,' what is that in our normal number system?

3) Do you know why and how you would use a loop?

4) Please check the box next to all the terms that are familiar to you. Do your best to describe the familiar terms.

☐ Variable:

☐ Method:

☐ Loop:

☐ Property:

5) What's a resistor? What symbol does it have in a circuit diagram?

Appendix J

Focus Group Interview Protocol

Camper Focus Group Protocol

G.A.M.E.S. Summer Camp

I am from the XXXX and I am conducting an evaluation of the G.A.M.E.S. Summer Camp. In order to conduct the evaluation, we need to obtain data through a number of activities such questionnaires; interviews.

I would like to ask you whether you agree to participate in the study by allowing us to interview you. If you agree participate in this study, the interview will take approximately 10minutes of your time.

The information that is obtained during the evaluation will be kept strictly confidential. Any sharing or publication of the research results will not identify you or any of the participants by name. A final written report of the evaluation results will be disseminated to the program personnel and funders and may also be used for publication in scholarly journals, or for presentations at various professional conferences.

Your participation in this evaluation is completely voluntary. You are free to withdraw your consent to be interviewed at any time and for any reason without penalty. During the interview, you will be able to interrupt it at any point and for any reason. You are also free to refuse to answer any questions you do not wish to answer. These decisions will have no effect on your future relationship with the school. Your participation in this study involves no anticipated risks beyond those that exist in everyday life.

Camper Interview Questions

1. What made you decide to attend this camp?
2. Are you having fun?
3. What are you enjoying about your experiences?
4. At school, what do you like about your sciences classes? Math classes?
5. At school, what do you dislike about your sciences classes? Math classes?
6. Here at camp, what do you like about your lessons and project?
7. Here at camp, what do you dislike about your lessons and project?

8. Do you think your friends would like to attend this camp? Why or why not?
9. Raise your hand if you plan on going to college. What are some majors that are attractive to you? Why?
10. Is this camp helping you learn what it is like to be an engineer?
11. Is there anything else you would like for us to know about how this experience has influenced your thoughts and feelings about being a scientist/engineer?
12. Is there anything else you would like for G.A.M.E.S. to know about how to make the camp better?

Appendix K

Parent Interview Protocol

GAMES Camp
Parent Survey

Thank you for helping us learn more about the G.A.M.E.S. summer camp! Your opinions and comments are important to us as we continue to improve the program in the future. Please share any additional comments with us on the back of this page.

Please respond to the following questions:

Your daughter had fun during the week. _____Agree _____ Disagree
Please explain.

Your daughter learned new things during the week. _____Agree _____ Disagree
Please explain.

Your daughter's interest in engineering was strengthened by the camp. ___Agree ___Disagree
Please explain.

Did she talk about G.A.M.E.S. to her friend(s) who didn't attend? _____ Yes _____ No
Please explain.

Would you recommend G.A.M.E.S. to other parents? _____ Yes _____ No
Why or why not?

How did you find out about the G.A.M.E.S Camp?

Is there anything you would like us to take into consideration for next year's camp?

Is there anything else you would like us to know about how this experience has impacted your daughter?

Optional: Your Name: _____
Your Daughter's Name: _____

Appendix L

Camp Counselor Interview Protocol

Counselor Focus Group/Interview Protocol
G.A.M.E.S. Summer Camp

We are from the Department of XXXX and we are conducting an evaluation of the G.A.M.E.S. Summer Camp. In order to conduct the evaluation, we need to obtain data through a number of activities such as camp observations; questionnaires; interviews with children, camp counselors, and some parents.

We would like to ask you whether you agree to participate in the study by allowing us to interview you. If you agree participate in this study, the interview will take approximately 10 minutes of your time.

The information that is obtained during the evaluation will be kept strictly confidential. Any sharing or publication of the research results will not identify you or any of the participants by name. A final written report of the evaluation results will be disseminated to the UIUC Women in Engineering personnel and may also be used for publication in scholarly journals, or for presentations at various professional conferences.

Your participation in this evaluation is completely voluntary. You are free to withdraw your consent to be interviewed at any time and for any reason without penalty. During the interview, you will be able to interrupt it at any point and for any reason. You are also free to refuse to answer any questions you do not wish to answer. These decisions will have no effect on your future relationship with the school. Your participation in this study involves no anticipated risks beyond those that exist in everyday life.

Counselor Focus Group/Interview Questions

1. Describe your students' experiences during the activities?

2. What did you expect your campers to learn during the week?

3. Do you think your campers' interests in engineering were strengthened or weakened by the camp? How?

4. What went well with your lessons and activities? What can be approved for next year?

5. Do you think the campers understood what was expected of them?

6. Has participating in this camp affected your social connections with your colleagues and faculty? If so, in what ways?

7. Has participating in this camp helped your matriculation through your degree program? If so, how?

8. Is there anything that you think needs to be improved about GAMES overall?

9. Is there anything you would like us to take into consideration for next year's camp? Is there anything you would like us to know about how this experience has impacted you?